DOCUMENT RESUME

ED 241 018 IR 050 626

AUTHOR	Lancaster, F. Wilfrid; Smith, Linda C.
TITLE	Compatibility Issues Affecting Information Systems and Services.
INSTITUTION	United Nations Educational, Scientific and Cultural Organization, Paris (France). General Information Programme.
REPORT NO	PGI-83/WS/23
PUB DATE	Sep 83
NOTE	228p.
PUB TYPE	Information Analyses (070) -- Viewpoints (120) -- Reference Materials - Bibliographies (131)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Case Studies; *Citations (References); Computer Programs; *Computers; Databases; Indexing; Information Dissemination; *Information Services; *Information Systems; International Programs; Online Systems; *Standards; Thesauri
IDENTIFIERS	*Compatibility (Information); Information Transfer

ABSTRACT

This UNISIST publication discusses issues related to the compatibility and standardization of bibliograpic records, index languages, software, hardware, and other information systems and services. Following an executive summary, definitions of terms, and other introductory material, existing information systems with common standards are briefly described. A section on compatibility of bibliographic records then discusses codes for bibliographic description, record formats, a common communication format, data exchange problems, bibliographic references, journal titles, script conversion, authority files, filing order, serials holdings data, and bibliographic record problems. Index language issues and products discussed include vocabulary compatibility, the intermediate lexicon, an integrated vocabulary approach, microthesauri, the macrovocabulary, algorithmic approaches, multilingual processing, other aspects of terminological compatibility, and index language problems. The section on software compatibility covers programs and programming languages, command languages, database management systems, machine assistance for searching, user-system interfaces, and software problems, while the hardware section discusses computers, input/output devices, information storage devices, character sets, information transmission, computer networks, machine-machine interfaces, and hardware problems. Compatibility of numeric databases and the management aspects of compatibility are discussed. Also presented are a series of case studies showing how eight organizations have achieved some level of compatibility in information exchange. A series of conclusions, a 566-item bibliography, and a list of abbreviations and acronyms conclude the publication. (ESR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED241018

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PGI-83/WS/23

Compatibility Issues Affecting Information Systems and Services

"PERMISSION TO REPRODUCE THIS MATERIAL IN **MICROFICHE ONLY** HAS BEEN GRANTED BY
C. Coudert-Schklowski

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR050626

General Information Programme and UNISIST

United Nations Educational,
Scientific and Cultural Organization

Paris,
September 1983

Reproduced by permission of
UNESCO

PGI-83/WS/23

Paris, September 1983

COMPATIBILITY ISSUES AFFECTING INFORMATION SYSTEMS AND SERVICES

prepared by

F. Wilfrid Lancaster

and

Linda C. Smith

General Information Programme and UNISIST

United Nations Educational,
Scientific and Cultural Organization

BEST COPY

This document is a
photographic reproduction
of the authors' text

Recommended Catalogue entry :

Lancaster, F. Wilfrid
Compatibility issues affecting information systems and services/prepared by F. Wilfrid Lancaster and Linda C. Smith /¯for the_7 General Information Programme and UNISIST .- Paris : Unesco, 1981 .- v, 209 p.,30 cm .- (PGI-83/WS/23)

I - Title
II - Smith, Linda C.
III - Unesco - General information Programme and UNISIST

© Unesco 1983

PREFACE

Unesco, through its Division of the General Information Programme and UNISIST, has been, for several years now, promoting the elaboration of guidelines, methods, norms and standards for information handling, as well as their application by information systems and services, in order to achieve greater compatibility and facilitate the exchange of information.

Since no fundamental study on the compatibility issues pertaining to information systems and services has so far been made available, and since the information community throughout the world is becoming more and more aware of the necessity of resource sharing through all types of cooperation and networking, it was felt that this major gap should be filled.

For this purpose, Unesco established, in December 1982, a contract with Professor F. Wilfrid Lancaster, Graduate School of Library and Information Science, University of Illinois, who undertook the present study in collaboration with Ms. Linda C. Smith, from the same School. They were assisted by Mr. Jaime Pontigo, Instituto de Investigaciones Electricas, Cuernavaca, Mexico, in the preparation of the section relating to record formats.

The facts and opinions expressed in the study are those of the authors and do not necessarily represent those of Unesco. In our view, the importance of the study lies in the fact that it will increase general awareness of those in charge of operational information systems and services or starting to establish new ones with regard to compatibility issues.

BEST COPY

CONTENTS

Page

List of illustrations i

Executive Summary. ii

0. Introduction 1
  0.1 Definitions 1
  0.2 The channels of information transfer 3
  0.3 Information services 6
  0.4 Primary publication 14
  0.5 The physical item 15
  0.6 The scope 15

1. Systems with common standards 17

2. Bibliographic records 19
  2.1 Codes for bibliographic description 19
  2.2 Record formats 22
  2.3 Toward a common communication format 27
  2.4 Some problems of data exchange 34
  2.5 Bibliographic references 37
  2.6 Journal titles 40
  2.7 Script conversion 41
  2.8 Authority files 43
  2.9 Filing order 48
  2.10 Holdings data for serials 48
  2.11 Some problems relating to bibliographic records 49

3. Index languages 50
  3.1 Vocabulary compatibility 52
  3.2 The intermediate lexicon 57
  3.3 Integrated vocabulary approach 60
  3.4 Microthesauri 62
  3.5 The macrovocabulary 64
  3.6 Algorithmic approaches 68
  3.7 Multilingual processing 69
  3.8 Other aspects of terminological compatibility 72
  3.9 Problems relating to index languages 74

4. Software 78
  4.1 Programs and programming languages 78
  4.2 Command languages 81
  4.3 Data base management systems 84
  4.4 Machine assistance for searching 87
  4.5 User-system interfaces 91
  4.6 Problems relating to software 98

5. Hardware 100
  5.1 Computers 101
  5.2 Input/output devices 103
  5.3 Information storage devices 105
  5.4 Character sets 108
  5.5 Information transmission 111
  5.6 Computer networks 116
  5.7 Machine-machine interfaces 120
  5.8 Problems relating to hardware 123

Page
6. Numeric data bases 126

7. Management aspects 131

8. Case studies 133
    8.1 Introduction 133
    8.2 The Regional Bibliographic Centre, SCOPAL 133
    8.3 DEVSIS 133
    8.4 TOXLINE and POPLINE 137
    8.5 The exchange of bibliographic data: a case study in geology. . . 142
    8.6 CCL: a case study in standardization of command languages. . . . 145
    8.7 iNET: a case study in systems interconnection 149
    8.8 CSIN: a case study in machine-assisted searching 151

9. Conclusions 156

References cited 166

Annex: list of abbreviations and acronyms 206

LIST OF ILLUSTRATIONS

		Page
Figure 1	The research communications "cycle"	5
Figure 2	The major components of a typical information service	7
Figure 3	Online searching of multiple data bases through a single computer facility.	12
Figure 4	Online searching of multiple data bases through multiple computer facilities	12
Figure 5	MARC-like record format	25
Figure 6	The user-system interface in online information retrieval . .	93
Figure 7	Alternative configurations for user-system interfaces	94-96
Figure 8	Possibilities for interworking: videotex and Euronet	118
Figure 9	Global internetworking	119
Figure 10	The seven-layer ISO reference model	120
Table 1	Number of translation programs needed in an international MARC network	29
Table 2	Different forms of bibliographic reference	39
Table 3	Term profiles for the broad category "heat transfer" in two vocabularies	67
Table 4	Data elements in the component data bases of TOXLINE	139
Table 5	The POPLINE unit record	141

EXECUTIVE SUMMARY

As applied to information services, compatibility is a property that governs the extent to which two or more organizations can communicate and exchange data. The more compatible are the operations of these organizations, the easier this communication and exchange will be. The achievement of compatibility is important because it is a pre-requisite to efficient co-operation, interconnection or resource sharing.

An obvious way for two organizations to achieve compatibility in some activity is for both to perform the activity according to the same standards. Thus, the work of agencies of standardization is of utmost importance to the field, as is the recognition and adoption of international standards by the information service community.

Not all standards promote compatibility, but only those that facilitate some form of exchange among organizations. Moreover, compatibility can usually be achieved by organizations operating under different standards, although with less ease and perhaps greater cost. One way is to directly convert an output of one centre (A) into a form suitable for use as input to a second centre (B) as, for example, when a computer program is written to convert the record format of A into a format that can be operated on by B's computer. Another way of achieving compatibility among organizations is through the establishment of some type of mediating device that acts as a kind of switching mechanism, converting an output of any participating centre into a form usable by any other participating centre. Thus, a "common communication format" can transform a processing format used by A into one usable by B (and C, D and so on), and vice versa, an "intermediate lexicon" can perform the same function for the indexing vocabularies of the organizations, and a switching language can achieve the same results for the command languages recognized by computers in these several organizations. Convertibility can thus be considered an important means for establishing compatibility.

Computer and telecommunication technologies greatly facilitate exchange of information among organizations. For this very reason, the automation of information services makes the achievement of compatibility more important than ever before. On the other hand, computers can themselves promote compatibility since they allow certain forms of conversion (e.g., of record formats) to be performed automatically.

Information services are very largely involved in manipulating records representing physical items (books, periodicals, periodical articles and other carriers of information) and it is in these records and the way they are handled, rather than in the physical items represented or the handling of the physical items, that compatibility is so important. This applies most particularly to records in machine-readable form.

The extent to which two information centres, A and B, can exchange data, given that both process by computer, depends on the answers to several questions, such as: are programs available that will allow B's computer to read A's records and to manipulate A's records in a useful way, can the records of A and B be integrated, can A's programs operate on B's computer, can A's computer communicate with B's, and can terminals within A communicate with B's computer? The degree to which such questions can be answered affirmatively depends largely on how much compatibility exists between the two organizations in digital encoding, record formats, hardware, software and telecommunications conventions.

Various compatibility issues directly affect the user of information services. Consider a user wishing to perform online searches of several data bases accessible through a single computer centre. Relevant compatibility questions include: can the user's terminal communicate with the computer (compatibility in telecommunications conventions implied), can all data bases be searched in the same way (implies compatibility among the record formats and a single command language), can a search strategy be transferred from one data base to another (implies compatibility of vocabulary among the data bases), are all results understandable to the user (implies compatibility among the bibliographic records printed out or displayed), can outputs be merged and duplicates eliminated (implies sufficient compatibility to be able to recognize duplicate citations), and can the search outputs be stored in machine-readable form by the user (implies compatibility between output medium and the user's storage medium)?

The situation of the user wishing to search data bases in, say, three different computer centres raises additional questions: can the user's terminal communicate with all three computers, and can a single command language be used (requires that all three centres adopt the same standard or the existence of some switching language)? Compatibility issues also affect the ability of two users to communicate with each other (i.e., confer or exchange mail electronically).

In the past 20 years, much has been achieved in securing compatibly among bibliographic records. An international standard for a format to permit the exchange of such records (ISO 2709) has been widely adopted. Based on this standard, the UNISIST Reference Manual is to provide a complete working manual for the international standardization of the form and content of computer-readable bibliographic descriptions prepared by abstracting and indexing services and others. The MARC exchange format and its derivatives, including UNIMARC, have also been widely adopted by the library community, establishing a significant level of compatibility among libraries. The new Common Communication Format is intended to satisfy the needs of both communities and one must have hope for its widespread acceptance.

The development by IFLA of international standard bibliographic descriptions for a wide variety of materials has contributed substantially to achieving compatibility in this area, while the International Standard Book Number (ISBN) and the International Standard Serial Number (ISSN) allow for the unique identification of these bibliographic items and thus facilitate the interlinking of different data bases. More progress needs to be made in standardizing conventions for the conversion of scripts, or in making the conventions more compatible, and in name authority control, especially at the international level.

Much less has been achieved in securing compatibility in the processing of component parts of works, especially journal articles, which is particularly unfortunate since these form the bulk of the currently published literature. Not even a unique identifier yet exists, greatly complicating the recognition of duplicate citations among machine-readable data bases. Great variation in bibliographic citation practices still exists among secondary publishers and among primary publishers, causing numerous problems for their users. It remains to be seen whether the new international standard bibliographic description for component parts offers any amelioration.

Many possible approaches to securing compatibility of index languages have been studied, including the merging of vocabularies, the establishment of an intermediate lexicon, and the use of macrovocabularies and microvocabularies. Nevertheless, while structural compatibility can be achieved by following ISO standards for thesaurus construction, it is doubtful if any significant progress can be claimed in achieving compatibility among vocabularies beyond this basic structural level. More progress has occurred in securing compatibility in the general area of terminological control (as exemplified by the work of INFOTERM) and in multilingual processing via machine translation techniques.

Information system designers and users can benefit from developments in hardware and software compatibility that originate in other application areas. Standards developed for programming languages, command languages, input/output devices, storage devices, character sets, telecommunications, and computer to computer links help to establish more uniform components from which to build information systems. However, new developments in technology frequently outpace the development of standards. For example, new technologies which one could potentially incorporate in information systems--such as videodisc, videotex, and local area networks--become new sources of incompatibility. At the same time, new developments in technology have made conversion an alternative to standardization as a means of achieving compatibility in many cases. Where no standards exist or where standards are felt to be too restrictive to satisfy local needs, mechanisms for conversion may still allow a given system to achieve compatibility with another system for purposes of sharing resources.

Many problems still exist. Significant barriers to further achievements are attributable to differing requirements in various segments of the information service community, due to national and cultural differences, research versus popular needs, and the varying levels of bibliographic detail needed by, for example, primary publishers, secondary publishers and libraries.

The "phasing" of standardization activities causes its own problems: international standards tend to emerge, in a sense, too late for the more progressive organizations who have already developed their own standards and may face considerable conversion costs were the new standards to be adopted in toto. Lack of completeness in various international standards also causes difficulties.

Further progress in achieving compatibility among information services in general will depend on the ability of the community to promote international standards or agreements sufficiently authoritative and flexible to replace the many national, regional and sectoral standards that now guide much of the work of this community.

Information services have much to gain by the adoption of international standards. This applies particularly to the developing countries, who have most need to share their own resources and to efficiently exploit the resources of the developed world. Nevertheless, information services must guard against adopting a standard that may, in fact, reduce the effectiveness of the service for the particular audience for which it is designed. For example, the precise needs of the users of some specialized information centre may necessitate the development of a highly specific, tailored vocabulary. Adoption of an existing, more general thesaurus might not be advisable in this situation.

Unesco is urged to do more in promoting compatibility among the various components of the information transfer cycle--primary publishers, secondary publishers, libraries and information centres of all types--so that processing efficiency is improved in a co-ordinated and integrated way throughout the cycle. Moreover, attention must be paid now to securing compatibility in the area of electronic publishing since it seems likely that future human communication will depend much more on electronic forms than on the traditional print-on-paper medium.

Compatibility does not always depend on standards; in some cases it can be reached by agreement on norms and methods of working. Consequently, Unesco has the ability to promote compatibility, among developed countries as well as the developing, even in activities for which standards are still lacking.

O. INTRODUCTION

0.1 *Definitions*

As commonly used in English, the word *compatible* refers to an interpersonal relationship. Two individuals are considered compatible if they are able to "exist together in harmony". *Compatibility* is not the same as *identity*. As Shilling/442 points out, two individuals may be compatible; they are certainly not identical.

In the information service arena, the meaning of the word does not vary much from its more general use. Different systems, system components or system activities are compatible if they can operate together in harmony (e.g., can communicate effectively or exchange records with a minimum of effort). In other words, compatibility implies the ability of two entities to work *in combination*. In Russian, in fact, "sovmestimost" (i.e., compatibility) is frequently defined as "the ability to be combined". More formally, Henderson et al/211 have stated that:

> Systems are considered to be *compatible* when the results of processing in one system are immediately and directly usable by other organizations having similar but not necessarily identical systems. Where the products of systems in collaborating organizations can be used interchangeably by those organizations without special efforts or conventions, the systems are compatible. (Page 6)

The report of the UNISIST feasibility study/498 puts essentially the same notion in somewhat different words:

> Compatibility: a quality of systems whose products can be used interchangeably, notwithstanding differences in notation, structure, physical carriers, etc., without any special conversion machinery. (Page 147)

The principal reason for seeking compatibility is to facilitate *co-operation* or *resource sharing* among organizations. Co-operation is defined by Henderson et al/211 as

> . . . collaborative effort or sharing between organizations in actual processing of materials for information systems or in exchange of the products of such processing. (Page 6)

The same authors draw a distinction between compatibility and *convertibility*. The difference is one of degree, relating to the ease with which some transformation or integration can be achieved. If two machine-readable data bases are digitally encoded according to the same conventions and if they use the same record format, the data bases can be considered compatible. This should mean that both can be manipulated by the same computer software without any further processing. If each data base uses a different record format, however, they cannot both be manipulated by the same software. A program must be written to convert the format of one data base to that of the other. These two data bases are convertible but not compatible. Convertibility can be considered a more general property than compatibility. Many outputs of information services can be converted one to another, can be interchanged or can be intermingled. The ease with which these operations can be achieved depends on the degree to which these outputs are compatible.

The UNISIST feasibility study/498 defined convertibility as:

> a quality of systems whose products can be made interchangeable through "conversion" programmes. (Page 147)

Writing specifically about the formats of machine-readable records, Lebowitz/280 distinguishes among identity, compatibility and congruence:

> Identical formats are those, the rules for whose formulation are identical or whose bibliographic records, though derived from different rules, cannot be distinguished. (Page 196)
>
> Congruent formats are those whose bibliographic records are not identical but can be made identical by computer processing with appropriate software. (Page 197)
>
> Compatible formats are those whose records can be used in systems together with records from other compatible formats. Compatibility is the property which permits records in various communication formats to be integrated into a single, usable file. (Page 197)

In this context, congruent is more or less synonymous with convertible. Lebowitz is quick to point out that congruence is not necessarily a reversible property. For example, the indexing records of one organization, using a specific vocabulary, may be automatically convertible to a much broader vocabulary used by a second organization. The reverse transformation, however, would not be possible automatically. To take a case in point, the specific STREPTOMYCIN in vocabulary A can be automatically converted to the broader ANTIBIOTICS in vocabulary B, but the documents indexed under this broad term in B cannot be mapped to the specific antibiotic terms used by A without human intellectual intervention (i.e., re-indexing).

Identity is a property that is objective and absolute. Two entities are either identical or they are not. But compatibility and convertibility are much more subjective and relative. We can refer to the degree to which two entities are convertible (or congruent) or the degree to which they are compatible.

Clearly, compatibility is closely related to standardization (French/165). The more two organizations adhere to common standards, the more likely it is that their activities and outputs will be compatible. In illustration of this, if two information centres construct thesauri that adhere closely to the standards of the International Organization for Standardization (ISO), the two thesauri will be more compatible (and, therefore, more easily merged or converted one to another) than would be true if two different standards were followed or if no standards at all had been adopted.

This report is not about standardization per se but about the various types and levels of compatibility that affect the efficiency of information transfer operations. Nevertheless, since compatibility and standardization are closely intertwined, it is not possible to divorce these topics completely. Standards applicable to the various compatibility issues will be referred to, as appropriate, throughout the text, although no attempt has been made to identify all standards of possible relevance*.

---

* Standards affecting information transfer processes have been comprehensively reviewed in the UNISIST Guide to Standards for Information Handling/496, the relevant ISO standards have been collected into a single publication/247, and a compilation of standard practices in the preparation of bibliographic records has been issued by IFLA/462.

It seems useful to clarify the relationship between compatibility and standardization. The adoption of common standards by several organizations is the most obvious way of achieving compatibility in the activities of these organizations. However, not all standards have significance from the viewpoint of compatibility, but only those affecting the exchange of objects or data among organizations. To take but one example, a standard for the size of microfiche has compatibility implications: it affects the ability of various organizations to exchange, store and use microfiche. A standard relating to life expectancy of microfiche, on the other hand, while important in its own right, has very little relevance to compatibility since it does not affect the ability of organizations to exchange the microfiche. Nevertheless, this is a rapidly changing environment and technological developments can soon make compatibility issues important in areas where they were previously unimportant. For example, standards for type faces had little relevance to compatibility issues until it became possible to read text automatically by optical character reader (OCR) devices.

Compatibility can also be achieved without standardization. For example, some procedure or device can be used to reconcile different practices as when one record format can be transformed, perhaps by computer program, to a second format. Standardization, then, is the obvious method of achieving compatibility but by no means the only one.

A major reason for seeking compatibility is to allow the interconnection of systems or networks, which, according to Baker/22, implies a reciprocal relationship: one system can both feed another and be fed by it (e.g., can give and receive records). Baker contends that, while compatibility is essential for interconnection, the achievement of compatibility does not in itself assure interconnection.

Compatibility is also related to consistency which, in this context, has to do with regularity and predictability in the representation of data. Consistency is more easily achieved (and more desirable) with some types of data than with others. Thus, it certainly seems desirable that libraries and other organizations be consistent in deciding the form in which the author's name should appear in a record. This type of consistency is not difficult to achieve if all organizations adopt the same standards (e.g., cataloguing codes). On the other hand, it is much more difficult to achieve consistency in deciding what subjects a publication deals with and in the selection of terms to represent this subject matter. Moreover, consistency here may be less desirable since different organizations may be interested in the same publication for quite different reasons. Standards promote consistency, both within organizations and among organizations. The more that organizations adhere to standards, the more consistency will exist throughout the information service community, the more compatible one system will be with another, and the more likely it is that various forms of resource sharing will be economically and practically viable.

0.2 The channels of information transfer

The most important channels by which the results of research and application activities are disseminated in science and other fields are depicted at the macrolevel in Figure 1. The box headed "User Community" includes two components: (1) those individuals who are involved in research and development and (2) those involved in applying the results of research and development. The communication problem represented in this diagram is that of disseminating the results and experience of research, development and application activities rapidly and efficiently to those individuals who need and can profit by this information.

As the diagram shows, various members of the "user community" report the results of their research and development activities or of their experiences in some field of application. These reports can be written or oral. Much of this information is disseminated in a completely informal way. Information is exchanged by individuals through correspondence and by conversations, either face-to-face or by telephone. Some of this information, when assimilated, stimulates new research or applications. These new projects generate their own reports which in turn are disseminated. The communication process represented in the diagram is thus a continuous and regenerative cycle. Science and other scholarly pursuits could not survive without an efficient communication cycle to support them.

The informal channels of communication are represented by the paths of one to nine and two to nine. The other paths, all leading eventually to the "assimilation" activity, represent the formal communication channels. In practice, however, the distinction between formal and informal communication is not as clear-cut as suggested in the diagram. Some channels combine both elements. An obvious example of this is the professional conference. Papers presented at such a conference, whether published or not, are formal communications, while the informal exchange through conversations in hotel lobbies, bars and restaurants is clearly informal. Moreover, formal communications may be disseminated through informal channels, as when a scientist mails a reprint of a periodical article to several of his professional colleagues.

As depicted in the diagram, many individuals and organizations play various roles in the distribution of information in written form. Writing does not in itself constitute communication. A message must be received and understood before communication takes place. Publishers of primary and secondary literature fulfill a primary distribution function as well as a publication function. Some primary publications and a few secondary ones are distributed directly to the user community through purchase and subscription by individuals. Much of this literature, however, reaches its users through secondary distribution functions performed by libraries and by a wide variety of information providers: information centres, information analysis centres, computer service centres, information consultants, and others. These entities play extremely important roles in the cycle. They have the prime responsibility for acquiring the published literature, for storing it (thus creating a permanent archive of human achievement), for organizing and controlling it, and for its secondary distribution. Their secondary distribution activities include all services provided: document delivery, literature searching, and reference services of all types, including those provided from machine-readable data bases. Another path in the communication cycle, and one that has assumed increasing importance in the last few years, is that by which information on ongoing research is disseminated through formal channels.

The communication cycle has not always looked exactly as depicted in the diagram. While informal communication channels in some form are as old as science itself, other channels are much more recent in origin. For all intents and purposes, primary publication dates back only to the invention of the printing press, but the major primary publication of science, the periodical, emerged only in the latter half of the seventeenth century, and secondary publications in science began to appear only in the nineteenth century. Indexes to ongoing research are a twentieth-century development. The configuration of the cycle, then, has changed over the years and can be expected to change further in the future. New communication channels emerge, others decline in importance and eventually disappear. As one example, the distribution of secondary publications directly to the user community is now almost nonexistent and there is evidence to suggest that the primary

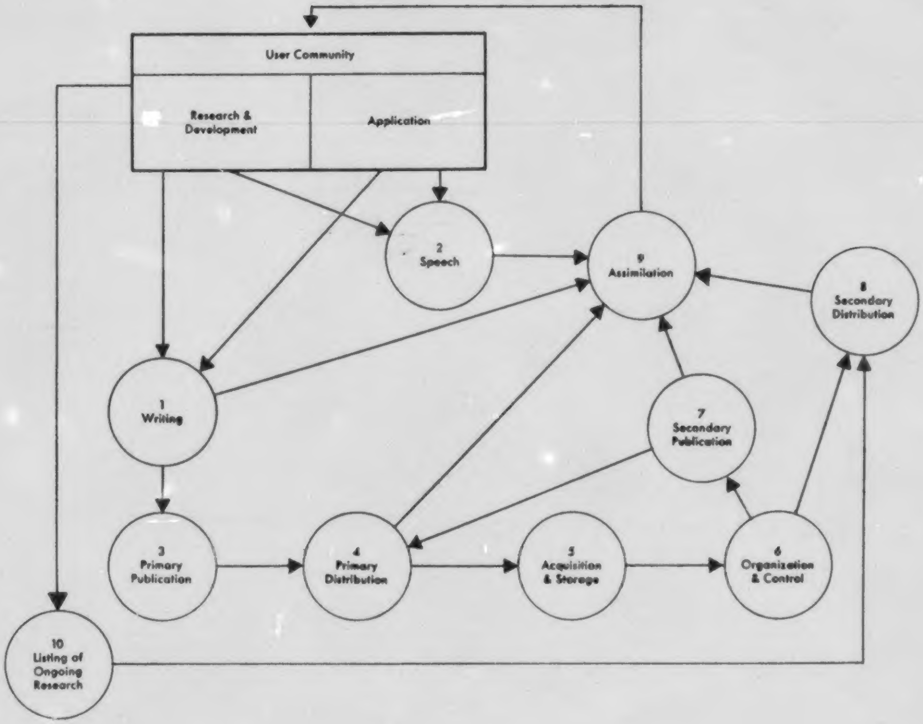

User Community
Research & Development
Application
2 Speech
9 Assimilation
8 Secondary Distribution
1 Writing
7 Secondary Publication
3 Primary Publication
4 Primary Distribution
5 Acquisition & Storage
6 Organization & Control
10 Listing of Ongoing Research

Figure 1. The research communications "cycle"

distribution of scholarly journals directly to the user community is declining relative to their secondary distribution through libraries.

As pointed out by Henderson et al/211, "the processes of handling, from the original generation through publication, dissemination, storage, selection, and ultimate use are typically interdependent". It follows, therefore, that information transfer within this cycle can be facilitated when certain activities and products are made compatible with each other. Conversely, lack of compatibility will obstruct efficient information handling throughout the cycle.

Dusoulier/131 explains that standardization and compatibility in information processing are of concern to three major groups: the producers (primary and secondary publishers), the distributors (libraries and information services), and the users of information. In the long-term, presumably, the major reason for seeking compatibility is to facilitate effective and cost-effective use of information products.

Unfortunately, wide variations in processing at various points in the cycle present barriers to efficient processing at other points. An obvious example is the wide variety of type faces used by publishers. While such variety may be aesthetically pleasing, it greatly complicates the subsequent handling of text by optical character reader (OCR) equipment in various large information centres.

This report will look at compatibility issues affecting all stages and channels of information transfer, but the principal focus will be those issues most relevant to the efficient operation of formal information services, namely nodes 5, 6, 7 and 8 of the network represented in Figure 1.

0.3 Information services

The major activities of most types of information service are presented in a somewhat simplified form in Figure 2. The system input consists of documents; that is, certain documents are acquired by the information centre. This implies the existence of selection criteria and policies, which, in turn, implies a detailed and accurate knowledge of the information needs of the community to be served. Once the documents are acquired, they need to be "organized and controlled" so that they can be identified and located in response to various types of user demand. Organization and control activities include classification, cataloguing, subject indexing, and abstracting. One important element is the physical description of the document (descriptive cataloguing) and the choice of access points (e.g., authors, titles) to allow this description to be found in catalogues or bibliographies.

As depicted, the subject indexing process involves two quite distinct intellectual steps: the "conceptual analysis" of a document and the "translation" of the conceptual analysis into a particular vocabulary. For efficient conceptual analysis the indexer needs both an understanding of what the document is about, that is, some comprehension of its subject matter, and a good knowledge of the needs of the users of the system.

The second step in the indexing process is the translation of the conceptual analysis into some vocabulary or "index language". In the majority of systems this involves the use of a "controlled vocabulary," a limited set of terms that must be used to represent the subject matter of documents. Such a vocabulary might be a list of subject headings, a classification scheme, a thesaurus, or simply a list of "approved" key words or phrases.

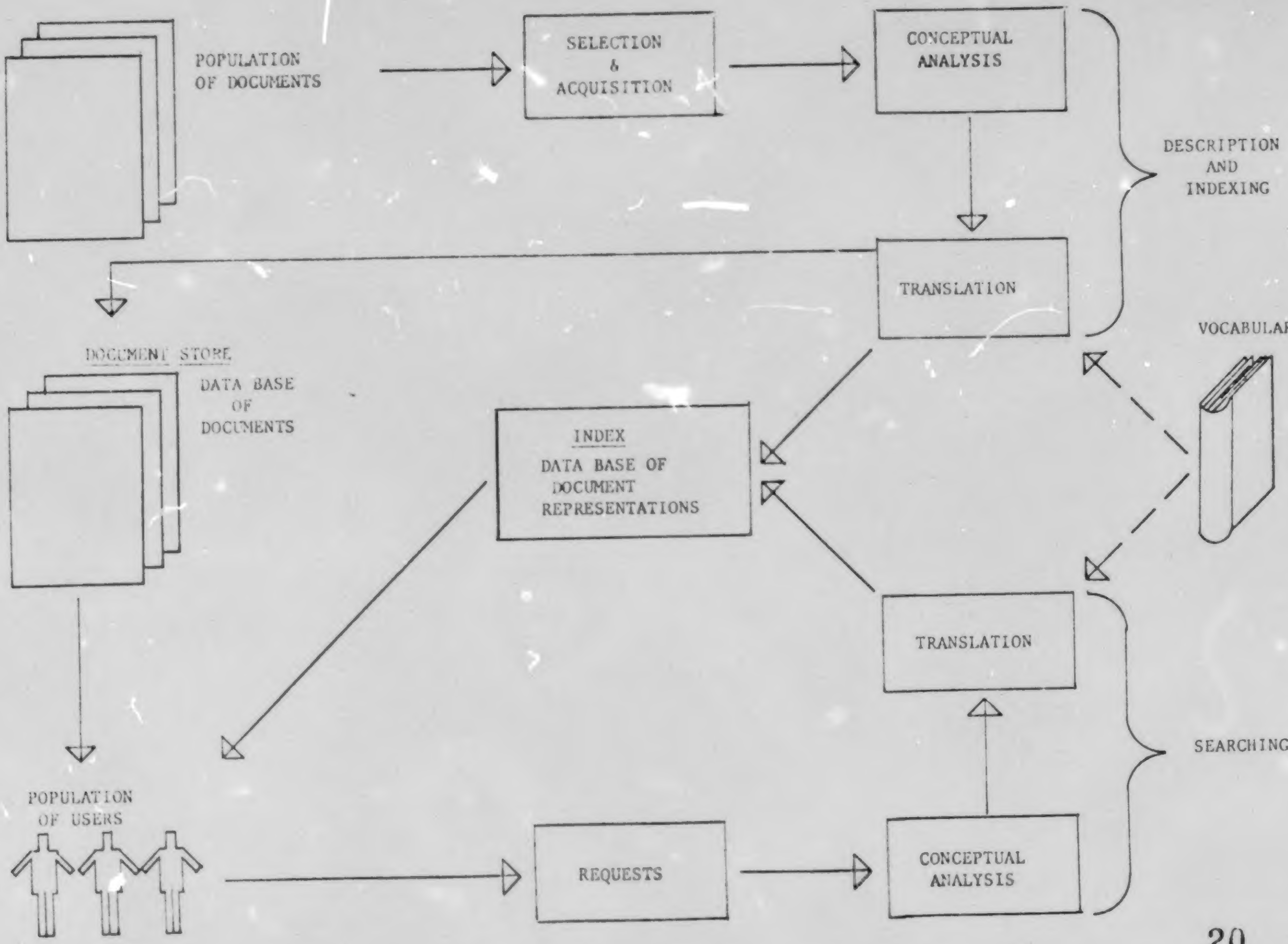

POPULATION OF DOCUMENTS
SELECTION & ACQUISITION
CONCEPTUAL ANALYSIS
DESCRIPTION AND INDEXING
TRANSLATION
DOCUMENT STORE
DATA BASE OF DOCUMENTS
VOCABULARY
INDEX
DATA BASE OF DOCUMENT REPRESENTATIONS
TRANSLATION
SEARCHING
POPULATION OF USERS
REQUESTS
CONCEPTUAL ANALYSIS

Figure 2. The major components of a typical information service

Once indexing has been completed, the documents go into some form of document store (document data base), and the bibliographic records representing them into a second data base, where they are organized in such a way that they can conveniently be searched in response to various types of subject, and other, requests. The data base of bibliographic records, or "document representations," may be as simple as a card file or an index in printed form. In a modern setting, however, it is more likely to be a machine-readable file on magnetic tape or disk. This data base may be regarded as the "index" to the document store. The "index" depicted in Figure 2 will consist of a bibliographic description of each document in the collection along with various access points to make this representation accessible, not only by terms describing its subject matter but also by such characteristics as author and title.

A bibliographic (document) data base is assumed in Figure 2. The steps will not change significantly if a different type of data base is involved (e.g., a numerical file or what, in Europe, tends to be referred to as a data bank). The contents must still be indexed to make them accessible but the "document store" of Figure 2 might be replaced by machine-readable tables of data (physical, chemical, numerical, statistical).

The steps involved at the output side of the service are very similar to those involved at input. The user population to be served submits various requests to the information centre, and members of the staff prepare search strategies for the requests. It is convenient to consider the preparation of search strategies as also involving the two steps of conceptual analysis and translation. The first step involves an analysis of the request to determine what the user is really looking for, and the second the translation of the conceptual analysis into the vocabulary of the system. The conceptual analysis of the request, translated into the language of the system, is the "search strategy," which may be regarded as a "request representation" in the same way that an indexing record may be regarded as a document representation. The only real difference between the two is that the former usually contains "logic," that is, a certain set of logical relationships among the index terms is specified; and the latter is usually without logic--logical relationships among index terms are not explicitly stated.

Once the search strategy has been prepared, it is "matched" in some way against the data base of document representations. This could involve a search of card files, printed indexes, microfilm, or magnetic tape or disk. Document representations that match the search strategy, that is, satisfy the logical requirements of the search, are retrieved from the data base and delivered to the requester. The process, which may be iterative, is completed when the requester is satisfied with the results of the search, which may, in some cases, mean that he is persuaded that nothing in the data base exactly meets his needs.

Some of the requests made to the information service will be for a specific document, for which author and/or title is known, rather than requests for information or documents on a particular subject. In the case of these known item requests, the request (see Figure 2) is made to the document store via author or title access points, or possibly by some other approach such as that of report number or patent number. The ability of the centre to supply needed documents may be referred to as its document delivery capability.

The major components of the service are evident in Figure 2. A typical information service may be considered to comprise six major subsystems:

1. The document selection subsystem.
2. The description and indexing subsystem.
3. The vocabulary subsystem.
4. The searching subsystem.
5. The subsystem of interaction between the user and the system (user-system interface).
6. The matching subsystem, that is, the subsystem that actually matches document representations against request representations.

It is clear that compatibility considerations affect the operations of this information service in several ways. To take a somewhat absurd example, if every technical report acquired were a completely different size, it would be extremely difficult to file them away in any efficient manner. In terms of storage requirements, these reports could be considered incompatible, although they may be completely compatible according to some other characteristic such as report numbering system.

It is in its relations with other information centres, and possibly with other types of organizations, that compatibility issues most obviously affect the operations of the service illustrated in Figure 2. Consider two information services, A and B, each one engaged in the range of activities shown in the diagram. Suppose the two organizations would like to co-operate with each other. For example, they may wish to exchange records so that each maintains an amalgamated data base of records compiled by each organization. If the data bases exist in card form, various levels of compatibility can be identified. Perhaps most obviously, the cards must be the same size, otherwise they cannot conveniently be interfiled. To facilitate use of the integrated file, it would certainly be desirable if the layout of the cards were sufficiently compatible that users could expect to find the same type of information in a particular location on each card. Finally, the integration will be most satisfactory if both organizations follow the same rules for descriptive cataloguing and use an identical or fully compatible vocabulary for representing the subject matter of documents. Here three levels of compatibility have been identified: physical compatibility, format compatibility, and intellectual (content) compatibility. Even if these two data bases are fully compatible at all three levels, integration is not necessarily a trivial task; nor is it cheap. Cards must be reproduced, exchanged at regular intervals, interfiled, duplicates eliminated or handled in some other way, and so on.

Suppose, on the other hand, that the two data bases are in machine-readable form, and that each information service has its own computer facility. In this case, the exchange of records is potentially easier and less expensive. Magnetic tapes can be exchanged each month between the two organizations at low cost (since the cost of duplicating a tape is minimal) and the actual integration of the two files, including elimination of duplicates, can be achieved through computer programs. Before this can happen, however, certain requirements must be met. In the first place, the records submitted by B to A (and vice versa) must be readable by A's computer. This requires that both sets of records must use identical codes for the digital representation of data. If this is true, A's computer will be able to read B's records and B's computer to read A's records. Being able to read a record, however, does not necessarily imply that anything useful can be done with it. A's computer programs will have been written to operate on records having a particular format. That is, the programs expect that each element of data will consistently appear in the same location in each record and that the data will always be presented in the same way. If B's record format is quite different, it will be incompatible with A's programs. A conversion program would be needed to translate B's record format into one that A's computer can recognize and work with. While this is certainly possible, lack of compatibility has hindered resource sharing between the two services and made this co-operation more expensive. As in the card file situation, it is highly desirable (for purposes of co-operation, at least) that A and B follow the same rules for descriptive cataloguing and that they use identical or compatible indexing vocabularies.

Other forms of co-operation may be sought between A and B. For example, A may already have computer programs for searching its data base and B may not. B would like to make use of A's programs. This may be possible if both organizations employ computers from the same manufacturer. If they do not, A's programs may be unusable on B's computer unless some type of conversion program is prepared, an expensive proposition.

Alternatively, let us assume that both A and B operate their information services through online computer facilities. B would like to access A's system and A would like to access B's. To achieve this, some equipment and communications compatibility will be needed. The computer used by A must be capable of communicating with the terminals used by B, and vice versa. If this technical compatibility exists, staff members of service A could get direct online access to the data base of service B. To use this data base, however, they must employ commands that B's computer has been programmed to recognize; they must use B's command language. Clearly, B's use of A's facilities, and A's use of B's, would be most easily achieved if both used the same command language. Otherwise, either (a) the staff of each service must learn two different command languages or (b) an interface must be prepared that will translate A's commands into B's and vice versa. Such an interface is not particularly easy to achieve and could be costly to construct.

Assuming a common command language, or the existence of some command language interface, a staff member of A could search B's data base in exactly the same way that he searches A's own. Of course, the vocabulary used to index B's data base may not be the same as that used to index A's. Again, either two vocabularies must be learned or one vocabulary must be "mapped" to the other. With a common command language and a common vocabulary (or, alternatively, interfaces to translate command languages and indexing vocabularies of one service to those of the other) an identical search strategy could be used to interrogate both data bases. This might be considered an optimum situation. The least optimum situation (from the viewpoint of efficiency at least) would be one in which each user would need to know two different command languages, be familiar with two different indexing vocabularies, and prepare a completely different strategy for interrogating each data base.

The situation discussed is simplified in that a bilateral relationship between two information services has been assumed. As the number of information services wishing to share resources, through some type of network, increases, the complexity of the situation is compounded.

Computers, telecommunications and machine-readable data have made it increasingly possible for information centres to share resources. A number of possibilities exist:

1. The merging of machine-readable data bases.
2. Conversion of the bibliographic records of one organization to those of another. This might include the formation of a tertiary data base (i.e., one in a specialized subject area formed by the extraction of records from several other data bases).

3. Establishing a network that permits any centre to get online access to the data bases of all other centres.

4. The situation as in 3 but made more sophisticated by allowing any user to search any data base by means of a single familiar tool: a single command language and, possibly, a single vocabulary of index terms.

The first of these requires compatibility in file organization and in the format of the records. The second requires, in addition, the ability to convert the index terms of one organization to those of others. The third, which is already common, requires only compatibility at the level of hardware and telecommunications protocols. The fourth, and most difficult, situation requires either the use of a common command language throughout the network or the development of an interface that will automatically convert one set of commands and search protocols to another. In addition, in the case of a network of centres in the same general subject area, it may require a similar interface to convert the index terms of one system to those of another.

A simple online search situation is shown in Figure 3. Here, a number of data bases are accessed through a single computer facility. The commands used to search the various data bases are likely to be constant since they are derived from software adopted by that particular computer centre. The different data bases, on the other hand, are all likely to have their own vocabularies of index terms, as well as differences in their indexing policies and protocols. Use might be facilitated were it possible to search several data bases in related subject areas through a single strategy, which implies the ability to convert one vocabulary to that of another. For the computer facility itself other compatibility issues will be important. For example, since all data bases will be processed by the same software, each data base must be made to "look alike" to the computer. This may mean that file structures and record formats be converted to a common structure/format.

A more complex but increasingly typical situation is illustrated in Figure 4. Here, a wide variety of data bases can be searched, through three different computer facilities, from the same terminal. The online user must now cope with three different command languages as well as the variations among indexing vocabularies. Again, it would be advantageous to learn only a single command language, and to have some network component to convert the commands of one language, as needed, to those of another. It would also be useful to have some network tool that would point to the data base or data bases most likely to satisfy a particular information need.

Some lessons can be learned from this discussion:

1. Compatibility is primarily related to resource sharing, in particular some form of exchange of bibliographic records or some means of common access to these records. Internal compatibility of processing within an organization will also be important but this is readily achieved by standardization or fiat. Inter-organization compatibility is a more difficult problem.

2. Computer and telecommunications technologies greatly facilitate resource sharing among information services. Indeed, some forms of co-operation, such as the exchange of very large data bases, would be virtually impossible in a non-automated environment.

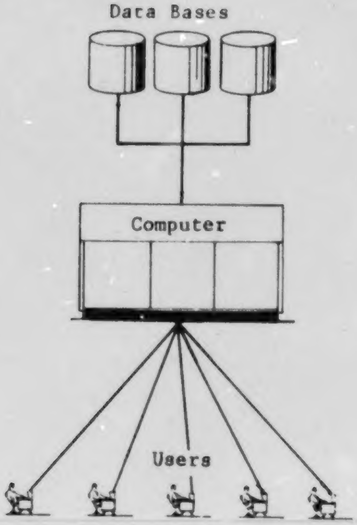

Data Bases
Computer
Users

Figure 3

Online searching of multiple
data bases through a single
computer facility.

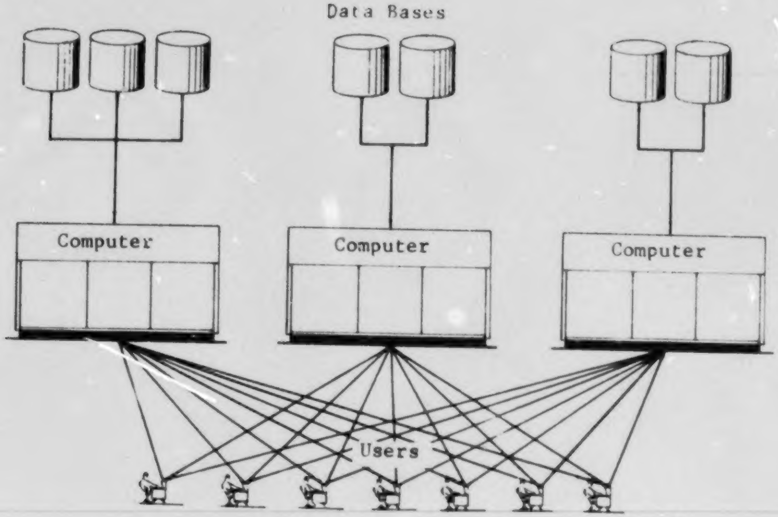

Data Bases
Computer
Computer
Computer
Users

Figure 4

Online searching of multiple data bases through
multiple computer facilities.

Reproduced from Niehoff/368 by permission of the copyright owner, John Wiley & Sons, Inc.

3. Because co-operation is facilitated, compatibility becomes increasingly important. Computers are less adaptable than humans. A human can compensate for lack of compatibility much more easily than a computer can (e.g., many can recognize roman numerals as easily as they can recognize arabic numbers).

4. In an automated environment, the absence of compatibility can add substantially to the costs of resource sharing.

It must be recognized, however, that adherence to standards, in order to promote compatibility, may have disadvantages as well as advantages. To facilitate co-operation with other organizations, an information service may have to make some compromises. The word "standard" can be considered an antonym for such words as "special" or "diverse". Since each information service has its own special community of users, with diverse needs, adherence to certain types of standards may actually reduce some aspects of effectiveness. For example, an information service may be ill-advised to change its indexing policies and protocols in order to conform more closely to those of other members of a network. If too many compromises must be made, the benefits of resource sharing may be completely outweighed by loss of local flexibility, autonomy or ability to tailor products to the needs of a specialized community. Further, when an information service has been operating independently for many years, it may be quite expensive to make the changes needed to bring its operations or outputs more in line with those of other organizations. In short, a particular centre must balance the requirements for co-operation with other organizations against the requirements associated with providing optimum information service to its own unique audience. It must also balance the costs of adhering to some standards against the possible benefits that might result from such adherence.

Vickery/514 has identified nine levels of organization of information by computer. The following tabulation shows Vickery's levels, in a sequence of increasing complexity, with one example given for each level:

Level	Example
1. Code for representation of binary data	EBCDIC
2. Character set	The Roman alphabet
3. Element set	A thesaurus of descriptors
4. Field	The descriptor field in a record
5. Record	A bibliographic reference
6. File	A list of report numbers
7. Data base	INSPEC
8. A system of data bases	DIALOG
9. A network of systems	EURONET

Clearly, compatibility issues can affect any one of these levels. For example: computers need to have binary data presented to them in a code they are designed to read (level 1); to allow processing by computer, a particular data element should appear in the same field in each record (level 4); data bases must be in the same format if they are to be handled by a common processing software (level 7); compatible telecommunications equipment and protocols are needed to interconnect systems within an online network (level 9), and so on.

Another way of looking at the subject is in terms of extent of the compatibility concerns. This can range from compatibility among various centres in a single organization to attempts to achieve levels of compatibility among an entire group of countries. The most notable example of the latter is the work toward establishing information systems compatibility among the countries of the Council for Mutual Economic Assistance (CMEA). This has been discussed at the general level by Kodola and Skripkin/268; more specific problems or applications have been described by Volny/518, Syrový/471, Haman/195, Krížiková/272 and Jedlička et al/254, among others. Work to achieve compatibility among the CMEA countries is spearheaded by the International Centre for Scientific and Technical Information (ICSTI) in Moscow. This centre has published a series of "normative technical documents" to guide the development, within the CMEA countries, of an International Scientific and Technical Information System (ISTIS). These standards pertain to bibliographic description, communication formats, physical characteristics of various media (magnetic tapes, microforms), and other aspects.

0.4 Primary publication

Compatibility aspects of primary publishing (steps 1, 3 and 4 of Figure 1) are considered relevant to the subject of this report only to the extent that incompatibilities in primary publishing may affect subsequent processing of bibliographic items by information services of all types. As hinted at earlier, computer and telecommunications technologies, applied to primary publishing, raise important compatibility concerns where they did not previously exist or did so only to a very limited degree. This is because:

1. By means of photocomposition, many publications are now generated from a machine-readable file. The machine-readable version of the publication can then be used as input to other operations.

2. The development noted in 1 allows the possibility of some level of integration between primary and secondary publishing.

3. Some publications exist only in machine-readable form. Many of these can be regarded as "secondary" rather than primary publications although others are, in effect, electronic reference "books" of various kinds. It seems only a matter of time before other types of publications (including scholarly journals) are available only in machine-readable form. Electronic publications raise compatibility issues (e.g., the ability to transmit, process and display machine-readable text) that are generally avoided by their printed counterparts.

4. Optical character reader (OCR) devices permit printed pages to be "read" electronically and translated into a machine-readable form.

Important compatibility issues arise from the fact that machine-readable text can be manipulated by computer in many different ways. For example:

1. The ability of an information centre to work with machine-readable text from many different publishers implies some level of compatibility in the way this text is encoded.

2. The ability to process full text by computer makes it increasingly important that publishers adopt common standards relating to such things as bibliographic citation and transliteration.

3. The ability to translate text automatically or semi-automatically makes terminological standardization more significant than ever before.

4. The ability to "read" printed pages electronically makes it desirable to limit the number of distinct type faces in use.

Other examples also exist but these at least illustrate the situation.

Recent technological developments also tend to blur distinctions among the various elements in the communication cycle. The work of primary and secondary publishers is now more readily integrated. Online systems make it easier for authors to be their own publishers. New, informal publications emerge within online networks of various kinds. Libraries, information centres and networks are increasingly involved in publishing, particularly in creating new data bases in machine-readable form. At the same time, publishers, by making their resources accessible through online networks, interactive television and other electronic media, are fulfilling a role more commonly associated with libraries.

National and international standards do exist for many facets of the intellectual preparation, presentation, editing, printing and publication of documents of various kinds. These are considered outside the scope of this report. For reasons stated earlier, however, certain aspects of the production of publications in machine-readable form are regarded as within scope and these will be treated as appropriate throughout the text.

0.5 The physical item

The same caveats apply to physical aspects of the carriers of information. Information services are primarily concerned with one type of physical artifact: the publication. While many standards apply to physical aspects of print on paper publications (e.g., paper quality), most have no real compatibility implications. Moreover, where compatibility issues do arise (e.g., the size of publications certainly affects the ability of libraries to store them efficiently), the concerns raised are relatively trivial.

Publications in other than print-on-paper form (including microforms, films and, most particularly, machine-readable media) raise more important compatibility issues since such forms can only be used by means of various types of equipment. Nevertheless, the compatibility concerns raised by physical characteristics of these media (e.g., the size of microfiche, the width of magnetic tape) seem so obvious that we have chosen not to dwell upon them in this report.

0.6 The scope

The fact is that the compatibility issues of most concern to information services of various kinds relate, not to the physical item per se, but to representations of this physical item and to representations of the content of the item. That is, the concern is primarily with bibliographic records representing published and unpublished documents of all types and especially with bibliographic records in machine-readable form. Publications that are themselves in machine-readable form (on magnetic tape, magnetic disk, optical disk, or whatever) present some of the same compatibility problems that affect machine-readable representations of publications, as well as some new problems of their own. It is these compatibility issues, which have to do with the ability of organizations to exchange data and the ability to operate on the same data in several ways (e.g., in order to derive different products and services) that are the prime focus of this report.

1. SYSTEMS WITH COMMON STANDARDS

It is obvious that a single information system, however large, can achieve system-wide compatibility by requiring its various system components to adopt a single set of standards. To quote one example, the National Library of Medicine in the United States receives input from a number of information centres in other countries but requires that its own indexing policies be followed and that its own Medical Subject Headings be the standard for terminological control.

Other networks, such as that of the Educational Resources Information Center (ERIC), also in the United States, are able to require component centres to adopt a common set of standards covering a wide range of activities. Similar examples exist in the decentralized international systems, most notably INIS and AGRIS, existing within the United Nations. Such systems achieve compatibility in input from the various component centres (which may be national or regional centres) through:

1. Standardized subject headings or descriptors, controlled perhaps by a common thesaurus, and perhaps standardized codes for such things as names of countries.

2. Guidelines for bibliographic description, together perhaps with a common input sheet to standardize content and format of bibliographic records.

3. Specifications for magnetic tape formats, record formats and character sets (for those centres contributing records on magnetic tape).

4. OCR input instructions (for those centres submitting input in the form of pages to be read by optical character readers).

The rigorous application of such standards throughout a network allows the co-ordinating centre to accept input from a multitude of centres, possibly in a variety of compatible forms (e.g., typed input forms, OCR sheets, magnetic tape, and paper tape). All of this may then be merged by computer at the co-ordinating centre to form a data base, which may then be issued in printed and machine-readable form. For international systems such as INIS and AGRIS, application is facilitated by making the guidelines and standards available in a number of languages. Nevertheless, it is worth noting that standards, however well conceived, are open to various interpretations, especially in the area of cataloguing and subject representation. With any large network, careful training of personnel in the various centres is an essential element in the programme. Moreover, the co-ordinating centre must monitor the quality of cataloguing and classification performed by the various centres.

This type of information system is exemplified well by INIS (see, for example, Raghavendra Rao and Kamath/400 and Turkov and Chepkasov/490, AGRIS (Aubrac/17) and, to a lesser extent, other UN networks such as INFOTERRA. The fact that INIS and AGRIS have been made compatible in terms of records, tape formats and character sets allows both sets of records to be processed by related software at a common computer facility. In fact, the AGRIS Input Unit is co-located with the comparable unit of INIS at the International Atomic Energy Agency in Vienna.

INIS has produced a wide range of internal "standards" to guide the operations of the input centres. These cover descriptive cataloguing, subject indexing, transliteration, abstracting, and various other aspects of intellectual input. Authority lists are produced for titles of journals, for names of corporate authors and for report number prefixes. Other standards apply to codes and character sets and to paper tape and magnetic tape specifications and record formats.

A further advantage of the standardization that an international system such as AGRIS can achieve is the ability to generate compatible products at national or regional levels. Thus, such products as the *Indice Agrícola de América Latina y del Caribe* (generated by the Centro Interamericano de Documentación e Información Agrícola, Costa Rica) and *A Current Bibliography of Southeast Asian Agriculture Literature* (the Agricultural Information Bank for Asia) are, in content and presentation, fully compatible with the products of AGRIS.

But not all information services can establish standards in the way that AGRIS and INIS can. This is particularly true in the establishment of a network whose member centres have been in existence for some years and have long-established policies and procedures. Achieving some measure of conversion or reconciliation may be the only approach possible under such circumstances.

The systems that establish their own standards, of course, have their own problems. The standards must be kept up-to-date. The international systems, in particular, face problems in trying to devise standards acceptable to a wide range of centres, some sophisticated and some much less so, operating under the constraints of different languages, in some cases different scripts, and different cultural environments. Accommodating to all of these needs can be difficult and costly. Moreover, a centre participating in such a system may also need to cooperate with other centres, outside this system, that might follow quite different standards. Universal standards, rather than simply local or sectoral standards, are needed to solve this problem and allow compatibility *among* systems that have already achieved their own internal compatibility.

2. BIBLIOGRAPHIC RECORDS

The majority of information transfer systems, whether purely manual or fully automated, are concerned with the processing of records representing some "bibliographic item" or "document" (book, periodical, periodical article, report, or whatever). It is in the handling of these _bibliographic records_, rather than the handling of the physical artifacts they represent, that compatibility is so important.

The ability to process bibliographic records by computer has stimulated an unprecedented level of worldwide interest in the activity of bibliographic description and in means of achieving compatibility in such descriptions so that machine-readable records can be exchanged and integrated. The ultimate objective of current efforts is the elusive Universal Bibliographic Control (UBC), the purpose of which (Anderson/^{9}) is:

> . . . to make universally and promptly available, in a form which is internationally acceptable, basic bibliographic data on all publications issued in all countries.
>
> The concept of UBC presupposes the creation of a network made up of component national parts, each of which covers a wide range of publishing and library activities, all integrated at the international level to form the total system. (Page 11)

For our present purposes, a bibliographic record for a document can be broken down into four major components:

1. A description of the physical item itself (in terms of such elements as author, title, publisher and place of publication).

2. Selection of elements from the complete description (such as personal or corporate author) to become _access points_ through which the record can be retrieved.

3. A unique identifier for the document.

4. Terms representing the subject matter dealt with, which might be controlled terms, drawn from an authority such as a thesaurus, or words and phrases drawn from the document itself.

2.1 _Codes for bibliographic description_

The first two components have been standardized, to a greater or lesser extent, by means of _cataloguing codes_. In the English-speaking world (and in various other countries), the standard has been the _Anglo-American Cataloguing Rules_ (AACR), now in its second edition (AACR2 is the common abbreviation). Similar cataloguing codes have been produced in Germany (the RAK) and in the Soviet Union. The latter is widely used throughout the Soviet republics and is thus applied in a number of languages. In the last several years, IFLA has produced a series of "standard bibliographic descriptions" to cover the first component; these specify the data elements needed to completely describe a bibliographic item, the sequence in which they should be presented, and the punctuation conventions to be used with them. They are considered essentially "frameworks" for bibliographic description, flexible enough to allow an individual user to select those parts of the description needed to serve a particular purpose (Anderson/^{7}). The results of IFLA's efforts in this area are manifested in the _International Standard Bibliographic Description for Serials_, or

ISBD(S)/[246] and the International Standard Bibliographic Description for Monographic Publications, or ISBD(M)/[245] which, along with ISBDs for other types of material, formed the basis for rules for bibliographic description in AACR2, in the RAK, and in certain other codes.

The standard bibliographic descriptions are not exclusively concerned with machine-readable records and not directly with how such records should be formatted. They deal with the content of a record rather than with the representation of that content or with the content of the document described. Nevertheless, by standardizing content, they contribute significantly to the achievement of content compatibility in machine-readable records and thus promote the international exchange of bibliographic data. Their standard presentation can also assist in the understanding of bibliographic records even when these records appear in a language unfamiliar to a particular user. Moreover, by adopting a rather elaborate scheme of punctuation to signal the appearance of a particular data element, the ISBDs facilitate the handling of bibliographic records by computer.

Since the ISBDs for monographs and serials were produced, IFLA has published a more general bibliographic framework, capable of handling documents of all types (Verona/[512]). The ISBD(G), a general framework that lists all the elements needed to describe and identify bibliographic items and prescribes a sequence and punctuation for these elements, appeared in 1977/[244]. Other ISBDs have been introduced for music, for nonbook materials, for maps and for antiquarian materials and an ISBD(CP) for "analyticals" (component parts of works) is the most recent of the series. The ISBD texts are subject to a five-yearly review.

Despite being designed to facilitate international resource sharing, the ISBDs were basically oriented toward Roman scripts, and were not particularly useful, for example, for handling non-Roman Asian scripts which may run from right to left and have basically different conventions from the Roman scripts (e.g., punctuation and capitalization may not exist). The Library Association of Japan, in fact, introduced its own "JSBD". A review undertaken by IFLA, however, has produced evidence to suggest that the ISBDs are more hospitable to non-Roman scripts than previously thought; in its current five-year review, IFLA is working to improve the international validity of the texts (Anderson/[8]).

Quite apart from the Roman bias, the ISBDs have not been free from criticism. Vervliet/[513], for example, claims that several of the bibliographic fields, despite the punctuation conventions, cannot be recognized automatically (by OCR equipment).* Moreover, differences between ISBD(S) and the International Serials Data System (ISDS) have presented barriers to the consistent interfiling of entries from both sources in any kind of title display. However, recent work performed by IFLA has done much to reconcile differences between the ISBD(S) and the practices of ISDS. Heymans/[215] has criticized the punctuation conventions for their ambiguity and for the fact that untrained users will find the conventions confusing.

Some countries have developed their own standards for the unification of bibliographic description. The Soviet standard GOST 7.1-76, for example, is based on the ISBDs and is designed to embrace contents and layout of descriptions for different kinds of publications in various languages and to apply equally to entries in catalogues and bibliographies and to bibliographic references appearing in a variety of publications, including periodical articles (Gosin and Smirnova/[185]).

---

* In fact, however, this was not one of the original objectives in developing the ISBDs.

For many years no unique identifiers of bibliographic works (the third component of a bibliographic record as listed earlier) existed, except in a limited sense (e.g., the catalogue card numbers assigned by the U.S. Library of Congress served this purpose for items catalogued by that body). Two important steps toward Universal Bibliographic Control were the adoption of a system for uniquely identifying publications by International Standard Book Number (ISBN) or by International Standard Serial Number (ISSN). Since ISBNs or ISSNs are carried in IFLA's standard bibliographic descriptions, these descriptions serve to identify documents uniquely as well as describing them. To the extent that a bibliographic description of a document may include some subject indicator, such as classification number, the description also provides some indication of subject matter. For our purposes, however, representation of subject (the fourth component) will be treated quite separately. This component--terms representing a document's subject matter--is by far the most difficult to standardize or even to subject to some measure of compatibility. The problems are discussed in Chapter 3.

Unique identifiers do not yet exist for other bibliographic forms, although some projects to develop such identifiers (e.g., for sound recordings) are now underway. An American national standard (ANSI Z39.23-1974) offers a prescription for developing a technical report number (organizations following the standard should all arrive at report numbers that are both unique and compatible) and the National Technical Information Service, in the United States, acts as a "clearinghouse" for report numbers compiled according to the standard. An International Standard Technical Report Number Code (ISTRN) is now under consideration by ISO, based on the ANSI standard. It is proposed that central code authorities in each country (or possibly groups of countries) should monitor the system and co-ordinate the assignment of unique report codes.

The four components of the bibliographic record can, in fact, be ordered in a sequence of complexity in relation to the feasibility of standardization or ease of reconciliation, as follows: 3, 1, 2, 4. It is important to recognize, of course, that the first three components identify or describe the physical artifact (e.g., book) itself rather than attempting to represent its contents. The fourth component is entirely content-oriented.

Wilson/553 and, later, Wellisch/533 have suggested that, while descriptive control of the formal and physical characteristics of documents seems possible, exploitative control, i.e., mastery of their subject content, can only be achieved in part. Wellisch, in fact, draws upon the laws of cybernetics to show why complete exploitative control is impossible to achieve. By the same token, then, it is reasonable to expect that some measure of compatibility in descriptive control is achievable while compatibility in exploitative control is altogether more difficult.

The ISBDs are not designed to contribute directly to choice of access point (the second component in our list). Standardizing this aspect is also quite difficult. There are disagreements and different conventions within a single culture (e.g., the English-speaking world) and even greater inter-cultural differences. Nevertheless, a standard description of a document does make some contribution to solution of the problem.

As Hickey/217 points out:

> Bibliographic information not included in the descriptive process can hardly be selected as the basis for an access point; thus it is important, wherever possible, for the descriptive information to be identified

> and recorded in a standardized way so that all significant access points will be embedded in an appropriate context and can be uniformly identified for retrieval purposes. . . The ISBD . . . lays the groundwork for possible future agreement about the way in which access points are chosen and their form of representation in a bibliographic file. (Page 237)

While computer processing greatly facilitates the exchange of bibliographic data, the very application of computers also complicates the process of bibliographic description and increases the need for standardization. Cataloguing codes and standard bibliographic descriptions specify what elements should appear in a bibliographic record, what form they should take (e.g., in what form an author's name should appear), and in what sequence they should be presented. While this is enough to achieve consistency in catalogues in card or book form, it is not enough to achieve consistency or compatibility of bibliographic records in machine-readable form. Computers are less forgiving than humans. Not only must content be standardized, but the way the contents are represented, as well as the way they are physically structured in a *record format*, must be rigorously controlled. To achieve this control, various related record formats, as well as standards pertaining to these formats, exist.

2.2 *Record formats*

The *format* of a record refers to a method of organizing data so that each item of data can be unambiguously identified. In traditional card catalogues, data elements are readily identified visually, by such characteristics as position, sequence, and spacing. For a machine-readable record, however, one needs a more explicit way of determining the end of one data element and the beginning of another. In a machine format, the need for precision in determining the beginning and end of every data element has required the establishment of codes that can indicate to the machine what kind of element or part of the element it is and how many characters it occupies.

A machine format can consist of the following:

1. Data element: the lowest unit of information (e.g., an ISBN).

2. Field: a collection of elements (e.g., a field relating to "authorship" consisting, perhaps, of the data elements of a person's name, and possibly dates of birth and death). The fields are of two kinds:

   a. Variable fields, containing data elements with no predetermined length.
   b. Fixed fields, containing data elements that are always expressed in a predetermined number of characters.

3. Record: the complete collection of fields treated as a unit.

In our context, a record represents a particular bibliographic item (a particular book, serial, periodical article, or whatever).

A machine-readable *record format* will have three major components:

1. The *structure* of the record, which is the physical representation of data on the machine-readable medium (it has been likened to an empty container that provides a basic framework for the record and allows its contents to be "transported" from one centre to another (Rather/404)).

2. The content designators or labels, which are the means of identifying the data elements or providing additional information about each. Mackenzie-Owen/296 refers to these as codes used to represent the data elements.

3. The contents of the record, which are the data elements themselves, such as names of authors, titles and place of publication.

Each of these components is susceptible to some degree of standardization.

The basic standards for record structure derive from the MARC format developed by the U.S. Library of Congress. These standards include national standards such as ANSI Z39.2-1971 (American National Standards Institute), the German DIN 1506 and, more importantly, an international standard ISO 2709-1981, Format for Bibliographic Information Interchange on Magnetic Tape.

As Boiko and Polozhentseva/45 have pointed out, the ISO standard, as it now exists, covers the general structure of data recording and codifies the use of four different types of record structures. However, it:

> does not deal with semantic aspects - the meaning of labels, ways to form the bibliographic description, systems for coding data elements, etc. The standard. . . was not enough to solve the problems of data base file compatibility. (Page 31)

A group within ISO is now working on an international standard for content designators: Codes and Numbering Systems for the Representation of Contents of Bibliographic Data Elements. The standard for codes for the representation of names of countries is ISO 3166 (second edition, 1981) and a draft standard for languages (ISO R 639) is under revision.

A variety of international standards are directly concerned with data elements and their representation, such as standards relating to the transliteration of non-Roman alphabets. The so-called "Paris Principles"*, which are now reflected in most cataloguing codes, have helped promote standardization in the choice of what data element an entry will file under (i.e., choice of access points) and the form in which this data element is presented.

The most influential record format is undoubtedly MARC. MARC (Machine-Readable Cataloging) is a generic term referring to bibliographic data coded and recorded on magnetic tape for further manipulation by computer. The MARC format, on the other hand, refers to a specific set of conventions for representing an identified ensemble of bibliographic data. There have been two versions of the MARC format so far. MARC I evolved in 1966 as a format for a Machine-Readable Catalog Record, an instrument in the MARC Pilot System and the MARC Interim System that followed. In 1968 the MARC II format was put into use and the full MARC System began functioning. A MARC format exists for books, serials, maps, music, manuscripts, and films, and a format for files of data is under development.

The MARC I format had certain defects and inflexibilities that limited its general utility. The MARC II format was intended to be:

1. Hospitable to the representation of bibliographic information for all kinds of materials.

---

* The statement of principles drawn up at the International Conference on Cataloguing Principles held in Paris in 1961.

2. Flexible enough to be used as the basis for automating all kinds of bibliographic activities in a wide range of institutions.

3. Compatible with the wide range of computer configurations and programming languages that might be used in processing the records.

Given these requirements, it was not possible to design a format that would be equally efficient for all users. The alternative was to create a format for communication (i.e., interchange) and encourage each institution to use its own processing format that would best suit its particular needs. MARC II, a format for communication only, would then become a common denominator that would require only one translating program from MARC II to the internal processing format in order to be usable.*

A machine-readable format for a bibliographic record, such as that of MARC, is likely to consist of three major components (Figure 5): a leader, a directory, and the data fields themselves. As suggested in Figure 5, these components are closely interrelated. The leader provides general information about the record that follows: its length, the type of document represented, and possibly its relationship to other records in the file. The record directory can be likened to the contents page of a book; it shows which data elements are present in the record and how they are labeled (tagged). The remainder of the record is composed of the bibliographic data themselves, organized into coded fields and subfields.

The MARC format was endorsed by the former Committee on Scientific and Technical Information (COSATI), thus having a considerable impact on the automation of libraries and information centres in U.S. government agencies. The American Library Association (ALA) and the Association of Research Libraries have also supported its use, and it has been adopted by major library networks throughout the United States. The ALA continues its interest in MARC. One of ALA's committees, the Interdivisional Committee for the Representation in Machine-Readable Form of Bibliographic Information (MARBI) has become the channel by which the library community works with the Library of Congress in securing needed expansions or modifications of MARC formats.

The MARC format, both in its structure and its content designators, has caused a revolution in the creation of bibliographic data bases worldwide. The existence of many "MARCoid formats," i.e., formats based at least on the principles behind the MARC format, proves the impact that MARC has had. These MARCoid formats are either national or regional. Examples of the latter are MEKOFF-2 (CMEA Countries), INTERMARC (francophone countries), MARCAL (Latin America), and EUDISED (Council of Europe). National MARC formats include:

Australia	Australian MARC (ANB/MARC)
Belgium	MARC-BR
Brazil	CALCO
Canada	Canadian MARC; MINI-MARC
Denmark	Marc Formatet (DN/MARC)
France	MONOCLE

---

* An exchange format is a carrier of data from one system to another. Its most important attribute is that of providing a standard method of exchanging information from one system to another, while satisfying the diversity of users and systems. (Avram/[19])

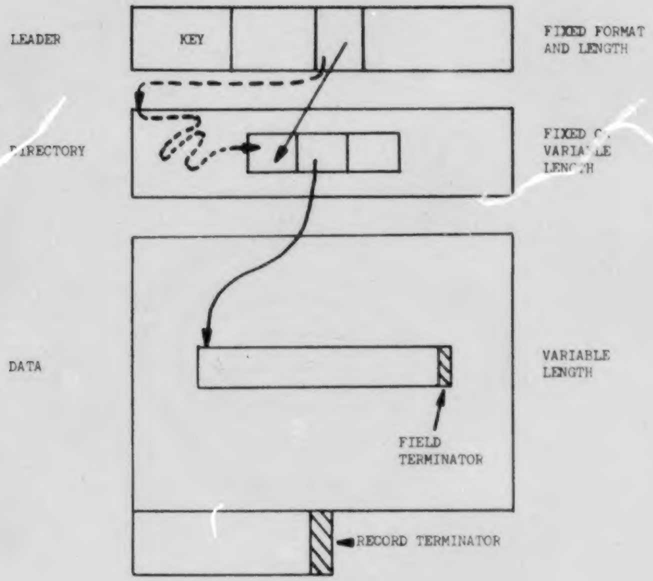

LEADER
KEY
FIXED FORMAT AND LENGTH
DIRECTORY
FIXED OR VARIABLE LENGTH
DATA
VARIABLE LENGTH
FIELD TERMINATOR
RECORD TERMINATOR

Figure 5. MARC-like record format.

From Somerfield/458 by permission of Aslib.

Germany (BRD)	MAB 1
India	Communication format used by NISSAT
Israel	MARCIS
Italy	ANNA
Japan	Japan MARC
Mexico	MARCAL-MEXICO
Netherlands	PICA
Norway	NORMARC
South Africa	SAMARC
Spain	IBERMARC
Sweden	LIBRIS
United Kingdom	BNB/MARC; U.K. MARC

In addition, the builders of various data bases (including INIS, AGRIS and INSPEC) have used the MARC structure as the basis for the creation of their own communication formats.

A further step toward achieving compatibility in bibliographic processing was taken in 1974 with the publication of the UNISIST Reference Manual for Machine-Readable Bibliographic Descriptions/406.* As described by Dierickx/118, the manual defines a carrier format structure, gives detailed specifications of the form and content of individual bibliographic data elements and defines the data element sets needed to describe particular types of documents. According to Buffet/63, the compilers of the RM made a concerted effort to observe the conventions of existing systems such as INIS and AGRIS. Hopkinson/224 claims that, not only does it describe an exchange format but, with a few additions, it could be made into a processing format. In addition, it may be used as a simplified cataloguing manual. He gives examples of organizations using the tool for each of these purposes, including Excerpta Medica, the CNRS Documentation Centre, the American Geological Institute and the Zoological Record. The manual, prepared by the UNISIST/ICSU-AB Working Group on Bibliographic Descriptions, was developed with the needs of indexing and abstracting services uppermost in mind. A particular feature of the RM is its ability to handle component parts of composite documents ("analyticals"), including periodical articles. The format was endorsed by the Association of Information and Dissemination Centers (ASIDIC) in the United States at its meeting of September 1978.

Avram/19, however, claims that the Reference Manual, by attempting to establish rules for data content in areas in which there is no international agreement (e.g., "authorship" or "intellectual responsibility"), precludes its [illegible] as the basis of an international exchange format. She points to basic conflicts between requirements of the Reference Manual and those of various cataloguing codes.

Dierickx/116, on the other hand, maintains that such criticism stems from a misunderstanding of the intentions behind the RM. The manual does not give rules governing such things as authorship, or even regulate the form in which names should appear, but only "prescribes a way of making the various subelements of a personal or corporate name separately identifiable so that any desired display can be generated automatically by computer program."

More recently, a reference manual for machine-readable descriptions of research projects and institutions has been adapted from the RM and issued as a Unesco publication/119.

---

* For convenience the abbreviation RM is used frequently in this report.

The UNISIST International Centre for Bibliographic Descriptions (UNIBID) was created in 1976 by the British Library under a contract with Unesco to operate within the framework of UNISIST. The objectives have been defined as:

1. To maintain the UNISIST Reference Manual (RM), to issue amendments and additions, and to promote its use, especially in making provision for suitable training.

2. To provide an information service on existing systems of bibliographic description.

3. To participate in the co-ordination of activities, including research and development, within the field of standardizing procedures of bibliographic description used by the information community.

4. To co-operate with other agencies or organizations operating within the General Information Programme in general and UNISIST in particular, such as those related to the IFLA/UBC Office.

An early activity of this agency was the organization of the International Symposium on Machine-Readable Bibliographic Exchange Formats, held in Taormina, Sicily in April 1978/120. This symposium provided the impetus for ongoing work on a Common Communication Format (CCF). Plans have been submitted to Unesco to develop portable software for a comprehensive information processing package based on the RM (see section 4.3).

2.3 Toward a common communication format

The integration of data bases requires the normalization of record formats so that, as far as the computer is concerned, only a single format is being processed. This can be achieved by writing a conversion program to map the format of the records in a particular data base to some standard institutional (i.e., processing) format. This is commonly done in information centres processing several data bases produced by different agencies.

As Boiko and Polozhentseva/45 have pointed out:

> A format implements a complex set of semantic and formal decisions relating to the set of elements used to describe a document of a specific type. When creating programs to convert data files we must therefore analyze both structural and semantic characteristics of the description of a document and find ways to convert from one structure and semantics to another. The formats of the systems under study will be compatible if we can compile a conversion program that enables us to extract from the files the bibliographic and semantic characteristics needed for retrieval in a given system. (Page 31)
>
> . . . structural compatibility can be viewed on the level of the general structure of records (for example, four types of structure in accordance with the ISO 2709-73 standard) and on the level of the formation of labels, the syntax used to record data elements, the alphabets used, and the codes and conventional symbols. (Page 32)
>
> The semantics of the format consists of the rules a system uses for bibliographic description of documents, an approach to a typology, standards for coding bibliographic elements, and for representing the con-

> tent of the document text. Thus, in terms of semantics, format compatibility is a matter of the compatibilty of recording the bibliographic characteristics of a document and the compatibility of semantic characteristics. (Page 32)

They go on to explicate these components of compatibility in more detail and illustrate the problems by means of a feasibility study. The study investigates the compatibility of the INIS format and of the INSPEC format with the format used internally by the Scientific and Technical Information Centre for Thermophysical Properties of Pure Substances at the High Temperature Institute of the Acadamy of Sciences (USSR). All three formats conform to ISO 2709.

This paper is of great interest in pinpointing some of the problems involved in this type of conversion (e.g., data elements present in one data base but not in another and varying levels of detail used in bibliographic description). Despite these problems, however, they were able to prepare schemes and tables showing the correspondence of data elements in the formats of the three systems. These schemes and tables can then be used to write programs to convert INIS and INSPEC tapes to the internal institutional format.

A specific example of achieving format compatibility through use of the RM is given by Packer and Conning/[378]. The work was performed within DOCPAL (the Latin American Population Documentation System) at the United Nations Latin American Demographic Centre (CELADE). DOCPAL, in common with several other agencies within the UN, uses the ISIS information retrieval software. A DOCPAL bibliographic format was devised to operate with the ISIS software. To make the DOCPAL data base more accessible (in particular, to allow exchange with other population information servic[illegible] and other interested agencies) and to make it compatible with the DEVSIS data base, being created by the International Development Research Centre, as well as other related data bases within the UN, a computer program was written to convert from the DOCPAL/ISIS format (a processing format) to the exchange format of the RM. It was discovered that several additional fields had to be added to the RM format to meet DOCPAL's needs. The reverse mapping, RM to DOCPAL/ISIS, was not written but no particular problems were foreseen for such a conversion.

Packer and Conning identified two different approaches to conversion. One was to write a program specifically designed to accommodate DOCPAL's adaptation of the RM, programming internally all the rules of the specific conversions between the fields. This was rejected because it would lack more general interest and application and would not take account of future modifications or additions to the processing format or to the RM itself. The second option was to prepare a more generalized program capable of converting from any ISIS adaptation of the RM to the RM format. This, too, was rejected as being too complex and requiring too much programming effort. The compromise actually adopted was a program that will permit, for any data element where there is direct correspondence between ISIS fields and the RM, conversion between any RM-based ISIS format and the RM exchange format. For more complex data element conversions, however, a user of the program must write a special routine to solve each individual conversion. These special routines can be added without changing the basic logic of the program. It is also possible for a user to add fields not specifically designated in the RM.

The proliferation of record formats, albeit mostly derived from the same parent, has been cause for concern for some years. Duchesne/[127] pointed out that:

> The result will be a babel of national implementations which will quickly become unmanageable in a dynamic situation. (Page 44)

Rather and de la Garza/405 elaborate on the problem:

> . . . there are sixteen countries using MARC or MARCoid formats, and therein lies the problem. Some of these formats are very close to ours, some are very close to the British, some depart slightly from both of these, and there are one or two that are so individual that they are called MARCoid only by courtesy. (Page 165)

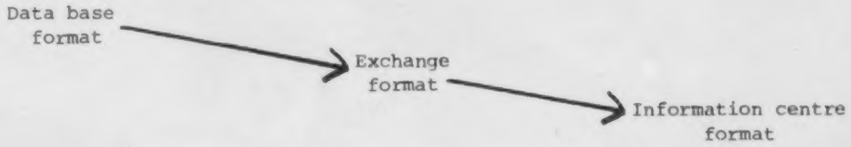
As a solution Duchesne proposed a truly international format to facilitate the exchange of bibliographic data in machine-readable form among national bibliographic agencies. He referred to this as SUPERMARC (1971). The major justification for SUPERMARC was that machine translation between formats (while possible) is uneconomical because of the number of translations that would need to be made in a truly international programme. Duchesne calculated the number of programs needed to cope with the translation problems using an international format, as opposed to the usage of different national formats, and arrived at the figures shown in Table 1.

Table 1

Number of translation programs needed in an international MARC network

Number of exchanging nations	Without SUPERMARC	With SUPERMARC	Saving
5	40	10	30
10	80	20	60
20	160	40	120
30	240	60	180

Clearly, a large centre processing many different data bases will be faced with a considerable amount of format conversion. If 50 data bases are processed, 50 conversions must be undertaken, assuming that all 50 use a somewhat different record format. An exchange format can reduce the conversion effort. If each data base format is made compatible with the exchange format, all the centre need do is to create a conversion program from the exchange format to its own internal format, as follows:



In this respect, the exchange format performs much the same function as a switching language or intermediate lexicon as discussed in section 3.2.

In studying the possibility of creating such an exchange format, Mackenzie-Owen/296 has characterized the format compatibility problem as follows:

> Standardization of the logical structure is a prerequisite for machine compatibility. However, the case is different with the implementation of that standard, i.e., the specification of data elements to be contained in the record, and their representation within the logical

> structure. These depend entirely on local requirements and on the rules for bibliographic description and entries in use. Such rules have not yet been standardized for all areas and all types of documents, even though recent developments, such as ISBD, have been a great step forward. There now exist various incompatible bibliographic formats (in the sense of specification and representation). (Pages 2-3)

He went on to compare, in some detail, the major MARC-based formats. MARC as used by the Library of Congress, MARC as used by the *British National Bibliography*, the Canadian MARC and INTERMARC employ essentially the same bibliographic elements and a common means of representation. All are based on AACR and have in common many field and subfield designations. Nevertheless, two major areas of incompatibility were still identifiable:

1. Codes and indicators used for the *processing* of information (e.g., sorting/filing instructions).

2. Detailed specification (e.g., data elements may be combined in a single field or subfield in one format but separated into distinct fields or subfields in another).

Moreover, some elements are specified in one format and not at all in another. It was determined that INTERMARC could cope with more complex bibliographic structures than the other three. Further differences were noted in other variants of MARC: MONOCLE, EUDISED and MAB 1.

Given that the formats have the same logical structure (as defined in ISO 2709), format translation (mapping) at the specification/representation level poses, according to Mackenzie-Owen, two essential questions:

1. What bibliographic elements are specified?

2. How are they represented within the logical structure?

He continues by pointing out two basic principles:

1. Only elements that are completely and unambiguously defined in *both* formats can be automatically mapped.

2. Given such complete and unambiguous definition, any difference in data element *representation* between the formats can be reconciled automatically.

This leads to the notion of *maxi-formats* and *mini-formats*. The former would contain the complete set of elements specified in the complete set of formats involved in some interchange environment, while the latter would contain only those elements that are *common* to all the formats. More precisely, Mackenzie-Owen would define a mini-format as one containing data elements *at the lowest level of specification*. Thus, place of publication would be a common data element even though, in some formats, it appears in a separate field while, in others, it shares a field with other data elements (e.g., publisher and date of publication).

The maxi-format and the mini-format both present problems for the interchange of information, in terms of what can be mapped automatically and what cannot. Mackenzie-Owen proposed a somewhat different approach:

> . . . combining aspects of both the maxi- and the mini-principle, and resulting in a general type of mini-format containing elements common to all formats, but retaining the possibility of exchange between formats of a higher degree of specification of elements not contained in the basic exchange format. (Page 8)

He developed the idea of a "universal format" which combines a "standard exchange format" (containing data elements common to all formats and therefore exchangeable among all agencies) and a "local format" (containing non-common data elements for local use only or exchangeable only with certain agencies). The common or standard data elements are those generated from the bibliographic item itself (i.e., elements that appear in an ISBD) plus additional elements standardized by the mutual agreement of all users. The nonstandard data elements would be all other elements, e.g., unusual elements not generally applicable or not of universal interest. Nevertheless, even some of these might be standardized among a small group of users by mutual agreement (Mackenzie-Owen refers to these as sub-standard).

Duchesne's original proposal, and some of Mackenzie-Owen's thinking, can be considered to be partly satisfied by the appearance of UNIMARC/495, a machine-readable format compatible with the ISBD, and giving specifications for the tags, indicators and subfield codes to be assigned to bibliographic records for monographs, serials, cartographic materials, films, music and sound recordings. It was therefore designed to serve the library community rather than the producers of more specialized data bases (e.g., the indexing and abstracting services), and its main purpose was to facilitate exchange of data among the various national bibliographic agencies.

As described by Avram/19 and by Avram and McCallum/20, UNIMARC was conceived of as a kind of "common denominator" for exchange purposes. The core element of the format is a "descriptive block" that contains the data elements included in the ISBD. These are standard in both content and content designation since they are independent of cataloguing codes. Those data elements that reflect lack of universal agreement (e.g., the intellectual responsibility, or authorship, of a work, as well as the representation of its subject matter) are considered "national" and will appear in various forms in the record according to national norms. Agreement was reached, however, on the "functional blocks" into which these data elements fall and it was thus possible to impose standard content designators on them to allow them to be read and recognized by a recipient agency. Thus, the record becomes a kind of "flexible container" for bibliographic information.

Dierickx/116, however, claims that the entire concept of including an ISBD block in the UNIMARC format, as well as repeating other data elements not easily standardized, such as personal and corporate names, for free local formatting, indicates lack of appreciation for the capabilities of computer processing. Provided that all elements and components (subelements) are separately identified in the record, they can be entered only once and formatted by computer program into a variety of desired displays.

Be that as it may, UNIMARC is intended to satisfy Mackenzie-Owen's requirement, for an exchange format, of allowing for both standard and nonstandard data elements. UNIMARC is one step beyond the standard information carrier (the record structure of ISO 2709). It provides standardization in content designation for all elements; it establishes which data elements must be included (the ISBD elements) and, for these, specifies the form the data are to take; for all other elements, however, it does not attempt to standardize their content. Its principal limitations are its restrictions on the types of documents accommodated and its orientation towards cataloguing conventions of the English-speaking world.

Avram/19 suggests that, with the appearance of the ISBD(CP) for "analyticals" (parts of larger works), UNIMARC could be extended to meet the needs of the secondary (indexing and abstracting) services. Others strongly dispute this, maintaining that the secondary services will never be prepared to accept the "unnecessary" redundancy of UNIMARC and that the ISBD for component parts, now being formulated, is totally unsuited to the needs of these services.

The Library of Congress has completed specifications to convert its MARC records to UNIMARC format and plans to offer tapes in the UNIMARC format early in 1983. The national libraries of Australia, Canada, and France, the State Library of South Africa, the British Library, and the National Diet Library of Japan are among other agencies who claim to be committed to the use of UNIMARC. UNIMARC has also been adopted as the basis of a common format for the exchange of bibliographic data between the U.S.A. and the U.S.S.R. (Gove et al/186), and now forms the basis of various national formats (e.g., in South Africa, Hungary, Taiwan). The maintenance and further development of UNIMARC is at present the responsibility of IFLA. However, an International MARC office is now being established in Frankfurt by the Deutsche Bibliothek, and this office may eventually assume responsibility for UNIMARC (Hopkinson/223). Work is now underway on the preparation of an interpretative UNIMARC manual, which will offer guidelines as to how to convert records into UNIMARC; in particular, it is intended to deal with the problems of data elements whose definition is not sufficiently developed to ensure compatibility of data originating from different sources.

Despite the existence of a number of formats designed for exchange or communication purposes (UNIMARC, the UNISIST Reference Manual, MEKOFF-2, and others), work continues toward the development of a common communication format (CCF), one that will meet the needs of libraries and documentation centres of all types, as well as publishers of secondary services, will accommodate documents of all types, and will be free from cultural biases. To this end, an international symposium, sponsored by Unesco, was held in Taormina, Sicily, in 1978/120. Not all participants agreed that an all-purpose format was either attainable or desirable. Malinconico/303, for example, argued that the requirements of libraries (dealing with publications at the macrolevel) and of the indexing/abstracting services (dealing at the microlevel) are difficult, if not impossible, to reconcile. Moreover, the library community, having already made a very substantial investment in MARC or MARC-like records, would be reluctant to convert to a new communications format, especially if the advantages could only be judged as marginal. A "monolithic" structure, to meet the needs of libraries and secondary services, was not, in Malinconico's view, the correct solution; it would invariably involve compromises that would fail to satisfy anyone completely. What are needed are efficient ways of linking the microlevel (indexing/abstracting services) data bases and the macrolevel (library) data bases.

Similar doubts were expressed by Hitzeroth/221 based on a comparison of the German MAB-1 format (for libraries) and the German MADOK format (for documentation): a common international format is a good idea in principle but it would not be effective if the indexing and abstracting services were forced to use more complex cataloguing procedures which would bring them little or no benefit. Avram/19, while less negative, doubted the need for a completely new format since UNIMARC could be expanded to meet the needs of the secondary services. Rowlett et al/420 calculated that it would cost Chemical Abstracts Service about $300,000 to convert from its present distribution format to a new format, without any retrospective conversion of records.

Despite these disagreements and doubts, those attending the symposium were generally in favour of pursuing the common communication format further. Among their recommendations were the following:

1. That a common format be treated as a high priority item by ICSU-AB, IFLA, ISO, Unesco, UNIBID, and other interested parties.
2. That the record structure be as defined in ISO 2709 and the emphasis now be placed on the establishment of standards for content designators and data elements.
3. That ISO undertake, as a matter of high priority, the compilation of a data element directory, including elements needed by all participants in the information community.

Since the Taormina symposium, several meetings have been held of an Ad Hoc Group on the Establishment of a Common Communication Format. A directory of data elements, drawn from the major international exchange formats now in use, has been compiled. Five principles were identified as the basis for developing the CCF:

1. Structure to conform to ISO 2709.
2. The core record to consist of those data elements essential to bibliographic description, identified in a standard manner.
3. The core record will be augmented by additional relevant descriptive elements, amenable to international standardization, identified in a standard manner.
4. Data elements not amenable to international standardization (e.g., classification numbers, index terms) will be identified by a standard technique.
5. A standard technique will also be used for accommodating levels, relationships and links between bibliographic entities.

Agreement was reached on the list of essential data elements by October 1979. Some of these were found to be missing in the ISDS guidelines and in the RM. Inconsistencies between the proposed form of certain data elements and UNIMARC, MEKOFF-2 and the RM were also noted. It was recommended that these guidelines and formats should be revised so that they include all essential elements in a standard form.

The initial design of the CCF (not merely a new format but intended as a bridge between the major international exchange formats) is complete. A printed draft will be produced by Unesco in 1983 to allow for international review, testing and evaluation.

Work on the development of a standard record format for the exchange of archives data is of quite recent origin. In the United States, the National Information System Task Force of the Society of American Archivists (SAA) established a working group to design a standard format for archival and manuscripts records data to allow the exchange of data and the eventual establishment of a national information system for archives and manuscript repositories. The development of the standard was a co-operative effort involving the SAA and other interested parties, including the Library of Congress and the Research Libraries Group. The SAA standard *Data Elements Used in Archives, Manuscripts and Records Repository Information Systems* was issued in 1982, as was the *Format for Archives and Manuscript Materials*, a MARC format maintained jointly by SAA and the Library of Congress.

Or[illegible] result of the application of computers to bibliographic processing has been a [illegible]dual movement toward standardization of the *content* of bibliographic record[illegible] In the United States, for example, the Library of Congress is promoting a "nati[illegible] level bibliographic record". In effect, this prescribes the data element[illegible]ibrary should include in a machine-readable record if this record is to form [illegible]t of some national data base.

Machine-readable bibliographic records have been given considerable attention in the past decade. Nevertheless, the needs of printed bibliographies have not been ignored. In particular, IFLA has issued guidelines pertaining to national bibliographies and the work of national bibliographic agencies/[192].

2.4 Some problems of data exchange

An exchange format is designed for efficiency in the *transfer* of data. It would not necessarily serve as an optimum processing format for any agency because such internal processing efficiency is hardware and software-dependent. A common exchange format would involve both data base producers and data base users, since both would be required to write conversion programs to and/or from an internal processing format. While adoption of a common processing format seems, on the surface, to be a cost-effective step for the information transfer community at large, data base producers (such as the major indexing and abstracting services) may feel they have little to gain, especially if this means accepting compromises. Even those data base users who have already written conversion programs for many data bases (data base distribution format to internal processing format) may doubt its value. Wilmot/[551] reports a feeling among data base users that format conversion costs are a negligible component of their computer processing costs. Neither is it certain that an organization planning a new system will be overly concerned with achieving compatibility with some exchange format. Cadwallader/[68], for example, discussing the requirements for a record format to be adopted by the National Agricultural Library (NAL) in the United States, urged that it:

> Should be based primarily on systems considerations (system efficiency, storage requirements, software complexity, etc.) with minimum weight given to the standardization of formats for communication purposes. (Page 4)

Several reasons were given to back up this recommendation, including:

> (1) A general-purpose standard tends to be not well suited to any particular system; (2) A standard format is likely to be biased toward communications usage and, therefore, serial media (tape) or serial telegraphic transmission, rather than toward a format most suitable for random access media; (3) NAL is likely to need data elements not present in the standard format. Similarly, the standard format will contain some of no interest to NAL; (4) A general-purpose format will not be most efficient; it may require needless software complexity; (5) A general-purpose format must be flexible and responsive to requirements external to NAL, leading to format modifications that would cause costly re-formatting at NAL.

Nevertheless, he recommended that, once a suitable standard communication format were identified, NAL should design and implement conversion software for accepting input and producing output in the standard format. This software effort was judged to be small "in comparison with the advantages gained".

In short, an internal processing format should be designed, as Cadwallader suggests, to meet local needs as efficiently as possible. This need not conflict with exchange requirements as long as the internal format is readily convertible to a suitable exchange format.

Regrettably, while compatibility is undoubtedly desirable to promote various types of co-operation in bibliographic processing, the very striving for compatibility may create problems of its own. Three possible causes are:

1. Conflicts may exist. To achieve compatibility in one direction may reduce compatibility in another.

2. Adoption of an international standard, when introduced, may force an organization to accept some undesirable compromises, sacrificing some local requirements for the sake of agreement with a larger group of agencies.

3. Adoption of an international standard may put a substantial burden, in conversion effort or costs, on an agency with a long tradition of adherence to a different standard.

One example of the first situation was that encountered in the relationship between AGRIS and the U.S. National Agricultural Library (Badran et al/[21]). Although NAL was one of the "prime movers" behind AGRIS, AGRIS adopted conventions that were compatible with INIS (to allow AGRIS processing to be performed on INIS computer facilities) and with the need to simplify cataloguing and indexing procedures so that they could be acceptable to a large number of national input centres, including those in developing countries. As a result, AGRIS was not compatible with NAL in record format, cataloguing conventions, indexing procedures or index language, forcing NAL into the cost of "double input", one for its own data base and one for AGRIS participation, and putting a serious obstacle in the path of NAL's full participation in the AGRIS programme.

All three situations can be considered illustrated by the Deutsche Bibliothek's participation in the ISDS programme. Wolf/[556] reports that this has led to double cataloguing and classification of periodicals at the library since the German cataloguing rules (RAK) differ from the cataloguing requirements of ISDS (elsewhere, Hesselager/[214] has reported similar discrepancies between ISDS and AACR, as have Sauer/[431] and Chan/[74]). Moreover, the Deutsche Bibliothek's record format (related to the German data exchange format MAB-1) was not compatible with the MARC-based exchange format adopted by ISDS. Adoption of ISDS guidelines would not meet the library's needs and would create incompatibilities with other German libraries. Equally, MAB-1 could not be jettisoned since it established compatibility within the German library community.

The third situation could be exemplified by the Chemical Abstracts Service (Rowlett et al/[420]) estimate that it would cost them approximately $300,000 to go from their present distribution format to an internationally accepted common communication format.

A common communication format does not solve the problem of fundamental differences in cataloguing traditions and conventions in various cultures. Kohl/[270] illustrates this through examples of problems involved in transferring German bibliographic data to an agency in an English-speaking country. Record format conversion (e.g., from MAB-1 to UNIMARC) is relatively easy but incompatibilities between cataloguing codes (AACR and RAK, the *Regeln für alphabetische Katalogisierung*) create major problems. Discrepancies involve the attribution of authorship as well as the form in which an author's name appears. Subject headings, classification and geographic codes also differ and, despite the ISBD, so do elements of the bibliographic description in some instances.

Anderson[6] has pointed to other problems obstructing international exchange of bibliographic data, including that of transliteration and of inconsistencies in national rules for the filing of bibliographic entries.

The international library community, as represented by the various national libraries and bibliographic centres, seems closer to achieving compatibility in its processing than has so far been achieved by the publishers of indexing and abstracting services. Moreover, comparatively little has yet been achieved in securing compatibility between both communities.

Publishers of indexing and abstracting services have for some time recognized the value of standardizing the data elements appearing in their machine-readable records. For example, the National Federation of Science Abstracting and Indexing Services (now the National Federation of Abstracting and Indexing Services--NFAIS) in the United States issued a report on data element definitions in 1971[350]. This report distinguished between elements considered essential and those considered supplementary. The desirability of achieving closer "alignment" with the bibliographic practice of the library community, as exemplified by the MARC project, was affirmed.

Nevertheless, in her study of eight major producers of machine-readable data bases, Wainwright[520] discovered that, while some follow ISO 2709 and the UNISIST Reference Manual, others do not. Even among those who do follow international standards, a number of variations were noted. Her findings were discussed further by Gray[188].

Wainwright refers to four characteristics affecting the ability to exchange or merge data bases in machine-readable form:

1. The physical attributes of the tape (such as size, number of tracks and packing density)
2. The organization of data within a record (i.e., record structure)
3. The representation of data on the tapes (such as character sets, codes and tags)
4. The content of the records (determined by such things as cataloguing codes and conventions).

ISO 2709 deals primarily with 1 and 2, while the RM, to a greater or lesser extent, addresses 2, 3 and 4. Wainwright notes variations at all four levels among the publishers of secondary services, a finding later confirmed by Wilmot[552].

In a detailed analysis of the printed bibliographic records generated by three major data bases in the United States, Larson et al[278] discovered that 15 different data elements were used, five to describe the journal from which an article is derived and 10 to describe the article itself. Eight of the 15 were common to all three services, four were used by two only, and three were unique to a single service. They raise the question that, since the records are very similar, "why shouldn't each be altered slightly to make them fully compatible?".

Henderson[209, 210] has undertaken a study of compatibility among various agencies of the United States government in the cataloguing of technical reports. It was discovered that, while differences exist in the number of data elements employed, as well as their definition and rules for use, some 17 data elements are common to four major report processing agencies and could thus form a "core". A data element dictionary, showing which elements are used by which agencies, is being prepared to promote further standardization of cataloguing among the agencies.

Piternick[389] has produced considerable support for Malinconico[303]. In a detailed review of the differences in purpose between library cataloguing and the indexing/abstracting services, and the effect of these differences on bibliographic description, she was able to show that similar data elements may have a different significance in one context than they do in the other. In particular, she indicates significant differences between the two types of organization in their handling of corporate authors. Moreover, substantial differences exist among the various secondary services in the handling of this and other data elements. Wide differences among indexing/abstracting services were observed in the handling of titles (in translation and transliteration, for example, and in the fact that some may actually add to the title) and their handling of titles tends to differ considerably from standard library practice. In general, little uniformity in bibliographic description can be found among the indexing and abstracting services. A similar type of analysis was performed by Gutierrez and Rey[193]. In their work, they examined and tabulated the data elements present in printouts generated from 33 online data bases.

2.5 Bibliographic references

Through its long history of cataloguing codes and conventions, the library profession has tended toward consistency in the way that bibliographic works are described or referred to. Much less consistency has existed elsewhere. The indexing and abstracting services have presented considerable diversity in the form of bibliographic reference used to identify a book, periodical article or other document type. Publishers of the primary literature exhibit even greater variability, in some cases bordering on eccentricity. The highly-respected U.S. journal, Science, for example, still excludes the title in citing an article appearing in a periodical. As long as it is accurate, of course, a bibliographic reference following any convention may be perfectly adequate to uniquely identify some publication. Nevertheless, lack of consistency does present problems: an author, submitting papers to five different journals, may need to follow five different conventions of citation; exchange of bibliographic records in machine-readable form may be complicated by variations in the number and sequence of the bibliographic elements; and the construction of citation indexes will be greatly complicated (Garfield[170]).

Standards governing bibliographic references are not a new phenomenon; work can be traced back to 1938 and a British Standard (BS1629) was first issued in 1950. However, producers of data bases and other publishers have frequently sought internal consistency by following a "style manual", published or for internal use only. In some cases, the manual was the publisher's own; in others, the publisher would adopt some existing manual. In the United States, the most influential of such authorities include those of the Modern Language Association of America[337], the University of Chicago[306], Skillin and Gay[448], and the Library of Congress[502]. These manuals differ somewhat in what they include in a bibliographic reference, in what sequence, how much is capitalized, and so on. Krummel and Howell[273] identify another family of style manuals, reflecting the practice of science journals, which they refer to as the "scientific list" group (also known as "Harvard style").

While intended primarily for machine-readable records, the UNISIST Reference Manual did establish a recommended form of citation for use in such publications as abstracting journals and printed bibliographies. ISO 690-1975, Documentation - Bibliographical References, issued one year after the UNISIST Reference Manual, offered a compatible reference format. This standard is now being revised to cover two types of bibliographic reference: a basic level for references appended to a text and a more detailed level for published bibliographies. The American National Standard for bibliographic references/5 was issued in 1977, although work toward such a standard had begun in 1966. The ANSI standard, Z39.29, claims to be compatible with the UNISIST Reference Manual and with the ISBDs, although the ISBD conventions for punctuation and format were not adopted in toto. It allows for an abbreviated as well as a full form of reference.

Krummel and Howell/273 have discussed the strengths and weaknesses of the ANSI standard. The full and abbreviated reference formats of ANSI are compared in Table 2 with the formats specified in various style manuals. The ANSI standard, however, besides prescribing a reference format for books and periodical articles, addresses itself to the citation of 16 further forms of documents. The ANSI standard is now followed, in principle at least, by several of the major data base producers.

At their 1977 meeting in Vancouver, the International Steering Committee of Medical Editors (ISCME), representing 19 editors of biomedical journals in English, adopted a slightly revised version and incorporated it into its Uniform Requirements for Manuscripts Submitted to Biomedical Journals. Fifty-five organizations in the United States had approved the standard before its adoption in Vancouver. Several of these were organizations that had previously adhered to an internal style manual. The National Library of Medicine adopted the standard for Index Medicus in 1980 and a strong movement is afoot to persuade biomedical publishers in general to adopt it (Huth/233).

Wellisch/531 has focused on one aspect of the bibliographic citation, the capitalization of words in titles, pointing to the diversity of practice in journals and data bases. ANSI Z39.29 and the Library of Congress both adopt common orthographic conventions for English: only the first word, proper names and acronyms are capitalized, at least for titles of articles or papers (the ANSI standard allows an alternative for titles of monographs or serials). Unfortunately, this convention is not followed by many publishers of journals and of data bases. In fact, some data base services cling to complete capitalization despite the fact that this has been shown to reduce readability (Paterson and Tinker/384). Wellisch deplores this diversity of practice and raises the legitimate question:

> Is it really necessary or useful to do this now that bibliographic data can be generated at one focal point, transmitted electronically to one or more databases, displayed at will in any desired configuration on CRT screens, transformed into printouts and retransmitted, perhaps by using one's own home computer linked by telephone lines to a receiving word processor or similar device in the editorial office of a journal? (Page 214)

Authority	Citation for books	Citation for periodical articles
Modern Language Association/337	Freeman, Susan T., and Laura R. Walters, comps. *Europeanist Social Anthropologists in North America: a Directory*. A Special Publication of the American Anthropological Association, No. 5. Washington: American Anthropological Association, c1975.	Irwin, Louis N. "Fulfillment and Frustration: the Confessions of a Behavorial Biochemist." *Perspectives in Biology and Medicine*, 21, No. 4, (1978), 476-91.
University of Chicago/306	Freeman, Susan T., and Walters, Laura R., comps. *Europeanist Social Anthropologists in North America: A Directory*. Special Publication of the American Anthropological Association, no. 5. Washington: American Anthropological Association, c1975.	Irwin, Louis N. "Fulfillment and Frustration: the Confessions of a Behavioral Biochemist." *Perspectives in Biology and Medicine* 21 (1978): 476-91.
Skillin and Gay/448	Freeman, Susan T. and Laura R. Walters, comps. *Europeanist Social Anthropologists in North America: a Directory*. A Special Publication of the American Anthropological Association, no. 5. Washington: American Anthropological Association, c1975.	Irwin, Louis N. "Fulfillment and Frustration: the Confessions of a Behavioral Biochemist." *Perspectives in Biology and Medicine* 21 (1978): 476-91.
Library of Congress/502	Freeman, Susan T., *and* Laura R. Walters, *comps*. Europeanist social anthropologists in North America: a directory. Washington, American Anthropological Association, c1975. vi, 34p. (American Anthropological Association. Special publication, no. 5)	Irwin, Louis N. Fulfillment and frustration: the confessions of a behavioral biochemist. Perspectives in biology and medicine, v. 21, 1978: 476-491.
ANSI/5 full	Freeman, Susan T.; Laura R. Walters, comps. Europeanist social anthropologists in North America: a directory. Washington: American Anthropological Association; c1975. vi, 34p. (A Special Publication of the American Anthropological Association; no. 5)	Irwin, Louis N. Fulfillment and frustration: the confessions of a behavioral biochemist. Perspectives in Biology and Medicine. 1978 (21) 4: 476-491.
ANSI/5 abbreviated	Freeman, Susan T.; Laura R. Walters, comps. Europeanist social anthropologists in North America: a directory. Washington: American Anthropological Association; c1975.	Perspect. Biol. & Med. 21: 476-491; 1978.
Scientific list	Freeman, Susan T., and Laura R. Walters. 1975. Europeanist social anthropologists in North America: a directory. American Anthropological Association, Washington. vi, 34p. American Anthropological Association. Special Pub. no. 5.	Irwin, Louis N. 1978. Fulfillment and frustration: the confessions of a behavioral biochemist. Perspect. Biol. & Med. 21 (4): 476-491.

Table 2

Differing forms of bibliographic reference as recommended by various authorities. Reproduced from Krummel and Howell/273 by permission of University of Toronto Press.

2.6 Journal titles

One element of a bibliographic reference in which standardization is certainly desirable is the title of a journal or other form of serial publication. This is of concern both to libraries (who maintain, and may want to integrate or exchange, lists of serials holdings) and to publishers of indexing and abstracting services. Moreover, variations in practice between libraries and indexing/abstracting services are becoming more of a problem as the resources of both groups become accessible through the same online networks. Thus, it is already possible to retrieve a citation from some data base of abstracts and switch to an online union catalogue showing which libraries subscribe to the periodical cited. Clearly, this type of switching is hindered if the two data bases use somewhat different forms of the title.

Further complications arise when journals change titles (or, worse yet, split into several parts) and in the handling of titles in foreign languages (which may be intact in the original language, transliterated or translated).

The most difficult problem of all, however, is the fact that periodical titles are frequently abbreviated rather than given in full. Abbreviation is possible in the secondary data bases and even more probable in bibliographic references appearing in the primary literature. Unfortunately, different journals and different data base publishers observe different conventions in the abbreviation of journal titles and abbreviations used are frequently ambiguous (e.g., Aust. could refer to Austrian or Australian). The publishers of *Chemical Abstracts*, of BIOSIS and of the *Engineering Index* have achieved compatibility by merging their lists of journal titles and reconciling differences/[40]. Nevertheless, wide variations among data bases do exist (Williams and Lannom/[545]). For obvious reasons, it is highly desirable that a journal be cited in the same way in different data bases. It would also be desirable that journal publishers should settle on a common convention for journal title abbreviation and that this convention be the same as that used by the producers of secondary data bases.

Some useful principles for the abbreviation of journal titles were put forward by Rogers and Charen/[410] in 1962, and the early history of standardization efforts has been summarized by Charen and Gillespie/[75]. An *American National Standard for the Abbreviation of Titles of Periodicals* (ANSI Z39.5-1969) was first published in 1963. The standard has been influential. It was followed by the UNISIST/ICSU-AB Working Group on Bibliographic Descriptions in the compiliation of the first *International List of Periodical Title Word Abbreviations* in 1970/[241]. The International Center for the Registration of Serials, located in Paris, now has the responsibility for keeping this list of abbreviations up to date and for issuing regular supplements (Szilvassy/[472]). The ANSI standard also formed the basis for the ISO code, ISO 4/1972, for journal title abbreviations. As Larson et al/[278] point out, however, even adherence to standards does not guarantee identical citation. Some journals, for example, may carry a slightly different title on cover and title page.

Unique identifiers for titles of journals do exist. The American Society for Testing and Materials (ASTM) devised a list of six-character codes (called CODEN) for periodical titles. The responsibility for assigning these now rests with the Chemical Abstracts Service. More importantly, the International Serials Data System (ISDS) assigns a unique number (International Standard Serial Number or ISSN) to each serial as well as standardizing on a "key title" which, according to Rather and de la Garza/[405], "is designed to be the preferred verbal form in citing that particular serial publication." Unfortunately, the cataloguing requirements of the ISDS conflict with other cataloguing codes, including AACR and the German RAK, a point that has been discussed by Wolf/556, Hesselager/214, Sauer/431, and Chan/74. More seriously, Vervliet/513 maintains that some differences between ISDS and ISBD(S) prevent complete interfiling of title entries formulated according to the two conventions.

Under Unesco contract, a considerable amount of work has been performed by the IFLA International Office for UBC to reconcile differences between the ISBD(S) and the practices of ISDS. As a result, the ISBD(S) review group and the ISDS Technical Directors have agreed upon specific details which, when incorporated within the ISDS Manual and the new text of ISBD(S), will allow the ISDS to derive the data elements it needs from the ISBD(S) as used in various national bibliographies. This should solve the problems of those national centres that have been forced in the past to make two sets of records.

The ISDS, of course, is more than a numbering system. It is essentially an international data base of serial publications; a considerable level of effort is required to keep such a comprehensive data base current.

Devices such as the CODEN and ISSN, which are language-independent, serve to uniquely identify a serial publication but do not solve the problem of access and recognition in information systems since users of these systems are unlikely to remember the unique identifier of a title and are even less likely to recognize the identifier if they see it. Nevertheless, unique identifiers of this kind are extremely valuable in machine-based systems and can serve to link records in various types of data bases. For example, if a record, displayed online, shows ISSN as well as journal title, the latter allows the user to identify the title while the former could be used as an efficient point of entry to other types of files (e.g., a journal location tool). Alternatively, the ISSN can be used to link variant forms of journal titles (including abbreviations, translations and transliterations) as they appear in various machine-readable data bases, allowing unequivocal identification whichever version a user chooses for access. Compact identifiers such as CODEN and the ISSNs can also serve to reduce storage space in computer-based systems (Piternick/390).

2.7 Script conversion

Wellisch/534 has estimated that some 375,000 bibliographic items--books and journal articles--are produced in languages using non-Roman scripts: Russian (Cyrillic), Japanese, Chinese, Devanagari, Arabic, and so on. The transliteration of these scripts into the Roman script (romanization) and vice versa--Roman to non-Roman--presents great problems and is a great potential source of incompatibility. Users of online data bases will be aware of the problem. For example, a Russian name may begin with C according to one scheme of transliteration, with T in another (e.g., Chekhov or Tjechov). This complicates certain types of searches performed in more than one data base (whether printed or electronic), makes it more difficult to identify duplicate citations, and puts other obstacles in the way of efficient resource sharing (e.g., exchange of bibliographic records, merging of data bases, and even interlibrary lending). As Wellisch points out, some agencies use the ISO transliteration standards/243, some follow Library of Congress practice, some follow national standards, and some use local adaptations and mixtures of standards. Wellisch claims that these various schemes are not "unambiguously reversible."

The situation is aggravated in data bases which index, for example, a Russian journal in its original form as well as in the form of a cover-to-cover translation. If the data base producer uses a scheme of transliteration different from that of the translation agency, the same name may appear in two different forms in the same data base. Another problem arises when a name goes through more than one transformation. Thus, a name converted from Japanese into Cyrillic and then into Roman may be quite unrecognizable.

Although ISO standards on transliteration do exist, Wellisch suggests that complete adherence to an international standard may be impossible for technical, economic and (most importantly) sociolinguistic reasons, which include differences in pronunciation of certain Roman letters in various languages and differences in orthography, especially use (or non-use) of diacritical marks. The requirements of different countries or language groups are thus basically incompatible. A significant practical problem is simply the varying requirements imposed by, on the one hand, human processing and, on the other, machine processing. INIS, for example, while following ISO standards as much as possible, has found the need to modify these to avoid use of diacritical marks.

Wellisch believes that forcing use of an international standard is not the proper solution. National bibliographic data bases must carry names (and possibly full bibliographic entries) in the original script, as well as in the transliteration of this script followed by the national bibliographic agency, and these two versions need to be cross-referenced in the machine-readable files. If necessary, entries in different scripts can even be merged for printout or display. For example, the New York Public Library (Malinconico et al/304) merges Roman and non-Roman scripts in the Hebrew section of its catalogue. This, of course, presumes the existence of the necessary character sets to allow, say, a British library to carry names in Arabic, Japanese, Chinese, or whatever other scripts may be involved.

Critics of Wellisch point out that, for technical reasons, use of an original script is not always feasible or economically desirable. Moreover, it does not solve the problem of incompatibility among various existing transliteration standards.

For indexing and abstracting services, Wellisch advocates transliteration of names and the translation of titles into the language of the data base. He believes that transliteration of titles serves no useful purpose in this context. For transliteration of titles of periodicals, he advocates following the practice of the International Centre for the Registration of Serial Publications (the agency responsible for assigning ISSNs); i.e., to use the romanized form of title adopted by this agency.

The ISSN, of course, is a unique identifier that is independent of language or script. Consequently, it can be used to switch from one language or script to another. Thus, through ISSNs carried in machine-readable records, the title of a Japanese journal can be transliterated (or translated) into Cyrillic (or Russian), Roman (or English), or whatever. Clearly, many problems raised by different transliteration would be solved if a unique number were assigned to the name of each author (personal or corporate)--an International Standard Authority Number (ISAN)--as discussed in 2.8. Building and maintaining a file of ISANs for all the world's authors, producers of periodical articles as well as monographs, would be a Herculean task and one that, even if feasible, might not be economically justifiable.

The whole subject of script conversion is a complex one. Wellisch has written extensively on the subject (see, in particular, reference 532).

2.8 Authority files

So far in this chapter we have dealt with compatibility issues relating to (a) bibliographic records for documents and (b) manipulation of these via computer processing. Clearly, many different access points--by name of author, by title, by publisher, for example--may be used in various information systems in addition to access points reflecting the subject matter of documents. In fact, in traditional librarianship, the non-subject access points are often regarded as more important than the subject access points. The catalogue of a library can be considered to serve two major purposes: (1) to indicate whether or not the library owns a particular document and, if it does, show where the document is located, and (2) to indicate what documents the library has on a particular subject. The former can be referred to as a known-item search of the catalogue while the latter, obviously, is a subject search.

For the library catalogue, satisfying the known item requirement can be considered the more important of the two functions. Other tools--subject bibliographies, indexing and abstracting services--can be used to determine what exists on a particular subject but the catalogue of a library is the only tool that indicates unequivocally whether a particular item is owned or not.

The most efficient way of conducting a known item search would be by means of a completely specific and unambiguous document identifier, such as an ISBN. Unfortunately, this is not really practical, at least at the present time, because:

1. Not all documents have ISBNs (or ISSNs)

2. Very few catalogues allow access via the ISBN

3. It is rather unlikely that the user of a library catalogue will know the ISBN for any item sought.*

What the user is most likely to know is the title of the publication, the name of the author (authors), or both. Consequently, it makes sense to assume that the major non-subject access points to bibliographic records are titles and names of authors.

In traditional library practice, the main entry in a catalogue is usually the entry appearing under the name of the person (or organization) that appears to be chiefly responsible for the authorship of the document. In other cases (e.g., when no author is known or no one author is clearly paramount), the main entry may appear under title. The term main entry is actually used two ways in the literature of library science: (1) to refer to that one entry in the catalogue in which full information on the document may always be found (all others, i.e., added entries, may give shortened information only), and (2) to refer to the access point under which this entry is filed in the catalogue, usually the name of an author, less frequently a title. In a computer-based system the whole idea of a main entry is redundant: a complete bibliographic record ("unit record") is stored once and all access points provided--by subject, by author, by title, by language, or whatever--are merely pointers to this record.

---

* These situations are likely to persist unless authors, publishers, professors and others can be induced to include these unique identifiers in their bibliographic references.

For more than a hundred years, librarians have tried to achieve consistency in the selection of access points (subject and non-subject) for library catalogues. For the subject access points, consistency has been sought by the publication of prescribed lists (i.e., controlled vocabularies). Thus, if we are to represent the subject matter "birds", we select, from some list, the access point BIRDS or PASSAROS or 598.2, or whatever--the list itself provides the consistency.

For non-subject access points, however, consistency has been governed, not by lists of this kind, but by rules to be followed, i.e., by cataloguing codes. The cataloguing codes address three major problems:

1) How to describe a bibliographic item concisely and consistently (i.e., bibliographic description--the function of the ISBDs).

2) How to decide which access point to select as the "main entry" and which access points to use as added entries.

3) How to formulate these access points consistently.

The formulation of non-subject access points, then, differs from the formulation of subject access points. In the latter, some authority (such as a thesaurus or list of subject headings) actually lists all of the acceptable access points; an indexer selects from the list. Cataloguing codes, on the other hand, do not list acceptable access points but provide rules to help cataloguers decide what access points to use for a particular publication.

Within the catalogue of a single library, it is desirable that a name be entered in the form that users of that library are most likely to select. It is not always possible to predict this. In any case, it is highly desirable that alternative forms should also appear in the catalogue as cross-references. Most important of all, however, is that a name be entered consistently so that works by the same author (personal or corporate) are not scattered in various places in the catalogue. To ensure such consistency, it has been common practice for a library to maintain some kind of list of the access points previously used in cataloguing. Such tools are frequently referred to as authority lists or authority files. As Malinconico/302 has explained, they serve two inter-related purposes:

1) to establish a precedent for a name (or other access point) so that this name consistently appears in the same form throughout the catalogue, and

2) to provide, by precedent and example, guidance on the form to be used for a name not previously encountered so that this form is consistent with earlier decisions.

In point of fact, once such an authority file is established, it achieves consistency in non-subject access points in much the same way that a thesaurus or list of subject headings imposes consistency on subject access points, i.e., by providing a list of the access points approved for use.

In the past, when libraries maintained such tools, they did so without much concern for achieving compatibility with other libraries. Nevertheless, some compatibility was achieved, almost accidentally, by adhering to cataloguing codes and by accepting as an "authority" the practice of some respected or influential entity (such as the British National Bibliography or the National Union Catalog). Indeed, many libraries chose to abandon their own, internal authority files in favor of adopting some such source as a "standard".

As the exchange of bibliographic data in machine-readable form has become more feasible and more prevalent, however, interest in authority files has increased considerably. It is not easy to merge data bases that use different authorities for names of individuals or corporate bodies. Nor is it easy for a user to consult an online network in which the name of an individual or organization may be represented in many different ways.

Resource sharing among libraries, then, could be aided by achieving compatibility among authority lists, e.g., reconciling differences in practice among libraries exchanging records or cooperating in a network. The reconciliation of authority lists is similar conceptually to the reconciliation of subject vocabularies, although the problems are rather different.

Authority files in machine-readable form are not new. Franzmeier/164, for example, has described the establishment of such a file, for names of corporate authors, at the Berlin Staatsbibliothek in 1971.

The Washington Library Network, one of the major online networks within the United States, has developed an online authority control file to ensure consistency and quality control among member libraries contributing cataloguing records to the network (Calk/70). Names of authors (personal and corporate), names of meetings or conferences, and "uniform title" entries (e.g., form of entry for various parts, editions, translations, etc. of the Bible), as well as subject headings, are all controlled. In this case, then, compatibility among the various libraries in the network is achieved by requiring each library to follow precedents already established in the authority file. In creating a new entry for the authority file, a member library must conform to network policies and standards. The entries submitted are subjected to machine and human review procedures designed to ensure quality and consistency. An extremely useful feature is the ability to search the file under any word appearing in an entry. This feature helps to ensure consistency of input as well as facilitating a user's search for names for which only partial information is known.

Perhaps the most extensive authority file for names is that maintained by the Library of Congress (Andre/11). The Library's name authority records are now distributed in machine-readable form. Part of the file, covering about 400,000 names, can now be accessed online in the United States. The National Library of Medicine also maintains a name authority file in online form. Since early 1977, the Library of Congress has been involved in a Name Authority Co-operative (NACO) project. The NACO project decentralizes the compilation of the name authority file, with a number of libraries and other institutions contributing data. The contribution of one NACO participant, Northwestern University, has been described by Hill/220, who concluded that:

> high-quality, consistent authority work can be performed in a decentralized manner . . . (and) successful creation of a consistent national bibliographical database depends heavily on the successful and efficient sharing of authority data. (Page 331)

In the NACO project the Library of Congress maintains editorial control over the headings established and entered into its authority file by other libraries. Early in 1982 some 23 institutions were co-operating in this project and some 38,600 name authority records had been submitted.

In the United States, too, the Council on Library Resources is actively promoting the "linking" of several separate online networks. The ultimate goal would be a truly national bibliographic network giving any library user, wherever located, access to records representing the holdings of virtually all libraries in the country. The Council believes that an "integrated consistent authority file" will be an essential component of such a national network and has prepared a detailed requirements statement for a Name Authority File Service. The component systems of the network seem to believe that the mutual exchange of authority data should be the first application of an intersystem linkage, preceding the exchange of bibliographic records themselves. A project to achieve such a file was launched in 1979. The Library of Congress is one of the participants. Unfortunately, while the Library of Congress has issued a MARC format for authority records, this format appeared after several of the networks developed their own formats, complicating the linking and integration process (Battelle Columbus Laboratories/[30]).

Clement/[84] has discussed the problems involved in developing and maintaining authority files within an international network. She suggests that each national bibliographic agency should produce authority records for all access points used in its own national bibliography. This view was also put forward by Poncet/[394].

Gorman/[184] has spoken of the possibility of establishing an international authority file--similar, in principle, to an intermediate lexicon--to facilitate the interchange of bibliographic information. Through the use of numerical codes, an English form of name such as Society of Jesus could be translated to an equivalent in French (Compagnie de Jesus) or some other language. An IFLA study/[242], on the other hand, suggests that a single international authority file would be unwieldy and impractical. A preferred approach would be the compilation of compatible nationally-maintained authority files with cross-references to allow conversions to be made automatically from one form of name to another.

The problem was recently addressed by Heymans/[215], who proposes an "international standard control form" (ISCF). The ISCF would be the form of name established by the appropriate national bibliographic agency (e.g., the Library of Congress for U.S. authors and institutions). However, the ISCF would not be imposed internationally. Different countries would follow their own national conventions. The ISCF would be an added element in all authority files and bibliographic descriptions, serving as a link between various national files. In effect, Heymans' proposal is conceptually similar to Gorman's.

Heymans goes on to indicate the requirements for achievement of Universal Bibliographic Control without sacrificing national needs:

> the International Standard Control Form is unambiguously identified as such; other uniform headings and reference forms can be added by the generating or by the receiving library without reframing the whole content; every heading and every reference form is identified as to the type of name (i.e., real name, pseudonym, abbreviated name, earlier name, later name, etc.), its language, the romanization scheme used, etc. Instead of framing the complete authority file description on the basis of a chosen "main entry," the generating library could indicate the form most frequently used in the country of origin; all name elements, prefixes in surnames included, are coded to make automatic recognition and manipulation possible; elements to be used to distinguish homonyms are given in a special subfield, in the vernacular form at least with a translation into English or French. (Page 162)

An IFLA Working Group on an International Authority System has considered the possibility of developing an "international standard authority number" (an ISAN analogous to an ISBN or ISSN) to uniquely identify names of persons, corporate bodies, geographic areas and even subject headings. Such a device could compensate for national variations in cataloguing conventions, language or even script. A unique identifier of this type would allow, for example, a name to be converted by computer look-up from a form preferred in the United States to one acceptable, say, in an Arabic-speaking country. If a standard bibliographic record (produced by a national bibliographic agency) were to consist of (a) an ISBD, and (b) a list of all names, personal or corporate, associated with the work,* an organization receiving the record in machine-readable form could select names to be access points or, in fact, use all of them for this purpose. If each name also had an ISAN associated with it, the problems of "name authority" and even of diverse transliteration (see 2.7) could be solved.

Unfortunately, there seems little immediate prospect of achieving the financial and administrative support needed to establish and maintain such a system. The working group has therefore turned its attention to alternative approaches/191. It has produced guidelines for printed authority records as one step toward the development of an international authority data network for sharing authority files. Work is also underway on a standard UNIMARC record format for the exchange of authority data.

IFLA emphasizes the importance of national experience in determining forms of names for national authors. The third edition (1977) of IFLA's Names of Persons, together with the 1981 supplement, covers name practices in approximately 65 countries.

It should be pointed out that not everyone believes authority control to be necessary. In an online catalogue, searches can be performed on any part of a name and cross-references can link records in which variant forms of the same name have been used, bringing all to the attention of the catalogue users. Malinconico/301, however, has argued strongly against this on the grounds that a catalogue, in whatever form, must collocate all related entries; it is not enough to provide rapid and efficient access at the level of the single entry only. Koel/269 has suggested that more research and experience are needed before this controversy can be settled:

> It is tempting to contemplate, in this connection, what the impact on the user would be if we constructed a catalog without normalizing headings. Cataloging costs would obviously decrease dramatically. Such a catalog would not display all works by an author together if the author had used different forms or fullness of name in his or her works. Nor would all editions and translations of a work be displayed together if they had been published under titles different from that of the original work. How much would this affect searches for a known item? Probably very little or not at all. Perhaps we should normalize names of literary authors only. Again, research is needed to answer these and other related questions. (Page 222)

---

* One of the recommendations of the International Cataloguing-in-Publication Meeting, Ottawa, August 1982, was that a CIP record should comprise bibliographic description (following ISBD requirements), "cataloguing access points" and subject access points.

Since the cost of establishing a name authority has been estimated to be as much as 40% of the cost of original cataloguing in a large research library, online access to authority information could offer the possibility of significant cost savings to libraries participating in a network. These economic aspects have been discussed in a report from Battelle Columbus Laboratories/30, which also identifies possible approaches to linking the authority files of several agencies.

2.9 Filing order

The sorting of records in alphabetical order (whether these are names, subject headings, titles, or whatever), for printout or for online display, is not "as simple as ABC." Even in a very simple file, some sorting (filing) rules must be observed. For example, sorting "word by word" puts SMITH, WALTER before SMITHSONIAN while a "letter by letter" arrangement, which ignores punctuation and blank spaces, would reverse the sequence. Sequencing of entries, of course, gets more complicated as:

1. The size of the file increases.

2. Numbers and letters are both used as entry points.

3. Different scripts are interfiled (see Malinconico et al/304 and Wellisch/529, 536).

Different filing rules now exist, either as separate publications (e.g., British Standard 1749) or as part of national cataloguing codes (as in the case of the German RAK). The International Organization for Standardization is now working on some "principles of bibliographic filing", taking existing rules into account, to be published as an ISO standard, as well as a code of filing rules to be published as a technical report.

2.10 Holdings data for serials

For many years libraries have co-operated in producing catalogues representing the holdings of several libraries (i.e., union catalogues), thus facilitating interlibrary lending activities. A special type of union catalogue is the union catalogue of serials--showing for any particular serial title, which libraries own it and how much of it they own (i.e., their holdings).

To produce a machine-readable union catalogue (e.g., one that can be accessed online) requires that co-operating libraries submit their holdings data in a common or compatible form. Various levels of compatibility are involved, including compatibility of the machine-readable format itself, as well as compatibility in cataloguing conventions (most obviously in the form of title adopted for the serial).

In the United States, a standard for holdings data (ANSI Z39.42-1980) does exist and work is in progress to develop a MARC format (compatible with the standard) for the exchange of location and holdings data in machine-readable form. A similar British Standard (5999:1980) also exists. Work is to be undertaken by the IFLA Section on Serial Publications to examine existing standards on holdings data and to provide a draft to present to ISO/TC46 as the basis for developing an international standard.

2.11 Some problems relating to bibliographic records

It should be clear from the foregoing that the last decade has seen great progress in achieving compatibility in the processing of bibliographic records, notably their processing by computer. Nevertheless, many problems still remain unsolved or even unaddressed.

Perhaps most obviously, while the library community has achieved a high degree of compatibility in the processing of bibliographic records for monographs and serial publications, little has yet been achieved in securing compatibility in the processing of bibliographic records for component parts of works, especially journal articles, which in fact constitute the bulk of the world's literature. Even less has been achieved in securing compatibility between the library community and the secondary publishers in bibliographic description and in record formats. Some would argue that the differences between these groups are basically irreconcilable.

The transliteration problem is still with us. Several conventions are now in use and the problems of converting automatically from one transliteration convention to another have not been solved. Much work is going on toward achieving compatibility in name control but we are still a long way from achieving, at an international level, the automatic convertibility from one form of name to another that was envisaged by Gorman (see 2.8).

Least has been achieved, however, in establishing the types of consistency in primary publishing (e.g., in bibliographic references) that affect the efficiency of certain secondary publishers, as well as of the services provided by libraries and information centres of all types.

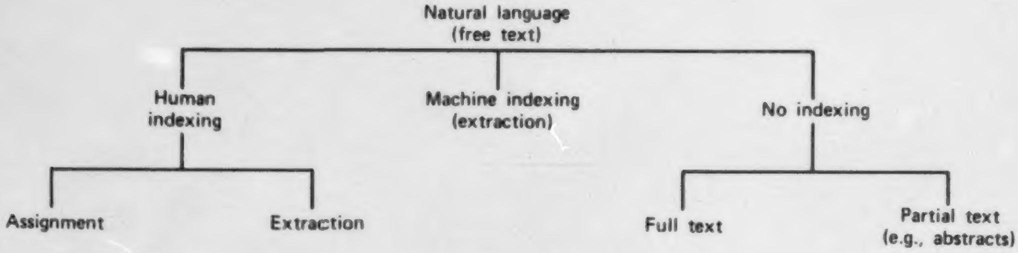
While unique identifiers exist for monographs (the ISBN) and serials (the ISSN), no such identifier has yet been implemented for journal articles or other component parts of works. Yet, such an identifier could significantly improve the efficiency of many information handling activities, including data base searching, interlibrary loan, and the interconnection of machine-readable files of various types. This is discussed further in Chapter 9.

In point of fact, the problems involved in securing useful levels of compatibility in the operations of the several components of the information transfer cycle (Figure 1)--between primary and secondary publishing operations, between publishing and other information services--seem to have received very little attention.

3. INDEX LANGUAGES

An index language is nothing more than a vocabulary of terms used to represent the subject matter of documents and the subject matter of information needs. Such vocabularies may conveniently be divided into (a) controlled vocabularies and (b) natural language.

In terms of possibilities, the natural language situation is somewhat more complicated than the controlled vocabulary situation. It may be depicted as follows:


Natural language (free text)
Human indexing
Machine indexing (extraction)
No indexing
Assignment
Extraction
Full text
Partial text (e.g., abstracts)

The term "natural language" itself means simply "the language of common discourse" (e.g., as used by authors in a particular subject field). In a natural language retrieval system, the subject matter of documents and of information needs is represented by an open-ended vocabulary of words/phrases commonly used in the subject field covered. The most significant term here is "open-ended". The antonym of "natural language" is "controlled vocabulary", i.e., some restricted set of terms that must be used by indexers and searchers. A controlled vocabulary in some subject field is some form of subset of the natural language of that field. It need be nothing more than this. That is, a controlled vocabulary could simply be a list of "acceptable" terms. More commonly, however, it will be given some "structure", as is the case with classification schemes, thesauri, and lists of subject headings. In the information retrieval context, then, a natural language system is one in which no control is imposed upon the vocabulary used to operate the system. "Free text" is sometimes used as a synonym for "natural language", although, strictly speaking, the term implies the use of words/phrases exactly as they occur in documents and thus can be considered more restrictive.

As shown in the diagram, a natural language system can be based on human indexing, machine indexing, or no indexing at all. When humans index using natural language they are likely to do so by extracting words and phrases from the text of documents, perhaps even by underlining them. This is indexing by extraction, or derivative indexing. The assignment of additional terms, not appearing in the text, could also occur. For example, an indexer might decide that the word "aerodynamics" should apply to a particular document even though it might not appear explicitly in the document itself. Natural language indexing by computer ("machine indexing" in the diagram) is always extraction indexing. While it is possible to program a computer to assign terms to documents, the limited set of terms that the computer could be programmed to assign would constitute, in effect, a controlled vocabulary.

In searching several data bases to satisfy a particular information need, it would seem desirable (at least on the surface) to use a single set of terms throughout. Merging of data bases would also be greatly facilitated if vocabularies were identical or, at least, compatible. Neelameghan/356 has put the case very concisely:

> To facilitate the integration and collaborative functioning of the information systems developed in different contexts, it would be helpful if the "languages" used for representation of subjects in the different systems are syntactically consistent, compatible with each other, and inter-convertible at a reasonable cost. (Page 166)

He goes on to point to two major approaches to achieving these goals:

> 1. To use the same or very nearly the same information storage and retrieval language in all the information systems, and
>
> 2. To use an intermediate language or switching language through or by which one moves from one information system to another. (Page 168)

The library community has long used common vocabularies, including such tools as the Dewey Decimal Classification (DDC), the Universal Decimal Classification (UDC) or the Library of Congress Classification (for arranging books and other physical objects) and standard lists of subject headings (for organization of entries in alphabetical subject catalogues). These tools are most applicable to the general, multi-topical library and are useful in achieving compatibility in the arrangement of many libraries of this type. It is relatively easy to accept a common vocabulary at this macrolevel. General libraries, almost by definition, involve compromises; they provide access to all subjects at a general level but fail to provide the detailed level of access needed by any one field of specialization. Even a general medical library may not be organized optimally for one whose specialty is neurology and even a neurology library may not be organized optimally for the specialist in epilepsy.

Adoption of a completely standardized vocabulary for all specialist fields, where detailed indexing of the contents of (say) science journals is involved, seems completely impractical. It is true that the UDC can and has been used for this detailed level of subject indexing but, as Shemakin and Kulik/440 have pointed out:

> The UDC, however, was created primarily as a hierarchical, precoordination system intended for manual retrieval. The results of the most recent research into the use of the UDC in the coordinate-indexing mode in automated IRS do not seem to offer much promise. (Page 36)

Moreover, the procedures that exist for keeping the UDC up-to-date are cumbersome and diffuse, contributing to reluctance in relying on this tool for microlevel indexing.

Information centres in the same subject field may be willing to use the same vocabulary, although this may mean the generation of bilingual or multilingual versions. The fact is, however, that little commonality of vocabularies does exist. Hence our present pre-occupation with achieving compatibility.

3.1 Vocabulary compatibility

Perhaps somewhat surprisingly, while controlled vocabularies tend to promote internal consistency within information systems, they also tend to reduce intersystem compatibility. Systems based on natural language are inherently more compatible than those using controlled vocabularies. Consider, for example, two data bases in the biological sciences, each one consisting of bibliographic citations plus abstracts in English. Given technical compatibility (in encoding conventions and record formats), these data bases are easily merged. More important, however, a single search strategy can be used to interrogate the merged file since the terminology used in one set of abstracts, being essentially the language of scientific discourse, should not differ substantially from that used in the other.

Suppose, on the other hand, that each data base is indexed by means of a different thesaurus. Given technical compatibility, the two files can again be merged. But they cannot be searched by means of an identical strategy, for a single concept might be quite differently represented in the two vocabularies.

Take the case of two information centres operating in the same subject field and wishing to share resources, either by exchanging data bases or by allowing each centre direct online access to the data base of the other. To simplify the situation, assume that both operate in English-speaking countries. For resource sharing (but not necessarily for other purposes) it would be desirable (a) that both use the same controlled vocabulary (which could, perhaps, be the Universal Decimal Classification) or (b) that both use natural language to represent the subject matter of documents.

If both use a different controlled vocabulary (e.g., different thesauri), the situation is more complicated. The data bases could still be merged, but each would need to be searched using its own set of terms. Moreover, it would be impossible to generate common outputs such as printed bibliographies arranged by means of identical descriptors.

For one reason or another, each centre wants to continue using its present vocabulary and is unwilling to accept the cost and inconvenience of changing to that of the other agency. Given this situation, and that centre A wishes to incorporate records from B and to produce common outputs, it will be necessary to convert ("map") B's vocabulary to A's. This means that, for each term in B's vocabulary, the nearest equivalent term in A's vocabulary must be determined and recorded. If the table of equivalents is stored in machine-readable form by A it would be possible to automatically convert B's records so that their subject matter is represented in the same terms as A's. The situation can be depicted diagrammatically as follows:

$$B \longrightarrow A$$

This will allow A to use B's records but it will not allow B to use A's. Mapping of B to A is not quite the same as mapping of A to B. Presumably, A will have some terms that do not appear in B and no equivalency relationships will be known for these terms if B is mapped to A but the reverse mapping is not performed. If each centre wishes to make use of the records of the other, reciprocal mapping must be carried out: A must be mapped to B and B to A, which can be represented as:

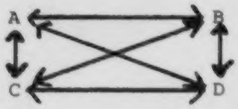
$$B \longleftrightarrow A$$

Standards, by promoting the structural compatibility of vocabularies, facilitate the conversion of one vocabulary to another. Thus, two thesauri following the ISO standards for thesaurus construction are likely to be more easily reconciled than two built on different principles. Moreover, such standards promote compatibility in a more general way: once he has become familiar with one thesaurus, it should be easy for a user of information services to convert to another thesaurus constructed according to the same conventions.

But structural compatibility is not the only consideration. Sokolov/456, for example, deals with the matter as follows:

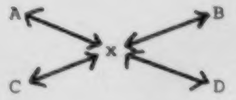
> Thesaurus compatibility means not only that the design and internal structure are uniform but also - and much more importantly - that they have the same working decisions concerning the formulation of keywords and descriptors, the creation of conditional equivalence classes, and the determination of paradigmatic relations. Ideal compatibility (obviously unattainable) would mean descriptor blocks could be mechanically exchanged between different thesauri. A practical approximation to this ideal would be the assimilation of parts of one thesaurus by another without much change in the paradigmatic structure of the latter. Only if thesauri are compatible can search patterns expressed in descriptor language be exchanged between different information systems. (Page 22)

The situation becomes increasingly complex the more centres that are to interchange records in some way. Let us say there are four such. Assuming that complete reciprocity is desired, this could be achieved by mapping each vocabulary to each other, which can be depicted as:

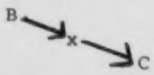


This is an expensive proposition for it requires 12 separate mapping operations.

There is, however, a possible alternative, namely the construction of a neutral "switching" language that can be used to convert from any one of the vocabularies to any one of the others:



Here, x represents the switching language (or "intermediate lexicon"). Equivalencies must be established between each component vocabulary and this neutral switching language. If this were done, a user of B could interrogate the data bases of A, C and D, as well as B's own, by means of an online network. He would use only B's vocabulary. Through the switching language, this would be converted to any one of the other vocabularies. To search C's data base via B's vocabulary, the path would be:



In other words, B's terms are converted to x and, through x as an intermediate lexicon, into C's terms.

With four centres, possible savings are not great since eight mappings (x to A, B, C and D reciprocally) are still necessary. Clearly, the more centres that wish to co-operate, the more likely it is that this approach will prove economical when compared with bilateral vocabulary to vocabulary mappings.

How easy it is to translate one vocabulary to another will depend upon several factors. The most obvious is the extent of the overlap in the subject matter covered. To use a somewhat absurd example, one would not expect a very satisfactory mapping between a vocabulary in engineering and one in medicine for the simple reason that they would have very few terms, or even conceptual areas, in common. Another important factor is specificity. If vocabulary A is large and highly specific, it may map rather easily to a smaller and less specific vocabulary, B. The reverse will not be true: mapping the very broad vocabulary to the highly specific one is likely to be less than satisfactory. A third factor is that of the degree of pre-coordination in the terms (Levy/287). The more alike the two vocabularies in this respect, the easier the conversion is likely to be. A vocabulary consisting mostly of single word descriptors is unlikely to map readily to one having many compound (pre-coordinated) descriptors, such as TURBOJET ENGINE NOISE. Finally, the extent to which the vocabularies are "structured", with hierarchical, synonymous and other relationships clearly displayed, will affect the ease with which a human is able to detect useful equivalencies between the two (see, for example, Wersig/537). Gopinath and Prasad/183 have produced a useful study of the extent of structural compatibility between thesauri and classification schemes.

Some of the problems involved in mapping one vocabulary to another have been explicated by Neville/366, who refers to the process as "vocabulary reconciliation". Neville recognizes various levels of correspondence between terms in two vocabularies, including:

1. Exact correspondence. This includes singular/plural variations. Thus, AIRFIELD and AIRFIELDS are considered identical. Also included are exactly synonymous terms in different languages. FLUGPLÄTZE, for example, is considered to exactly correspond to AIRFIELDS.

2. Synonymy. UNDERGROUND STRUCTURES, BURIED STRUCTURES and SUBSURFACE STRUCTURES can all be considered synonymous. Sometimes such synonyms are identified explicitly through cross references appearing in one of the thesauri. The reference Underground structures use SUBSURFACE STRUCTURES, for example, indicates that these terms are considered synonymous, at least by the compilers of one particular thesaurus.

3. Specific to broader term. The term SNOWDRIFTS, for example, appearing in vocabulary B, may need to be mapped to the more generic term SNOW in vocabulary A.

4. Term mapping at different levels of pre-coordination. For example, the term FROST PENETRATION in one vocabulary is considered equivalent to two terms, FROST and PENETRATION, in a second. In a more complex and less obvious example, the term STIFFNESS METHODS in A may be taken as equivalent to the terms STRUCTURAL ANALYSIS and DISPLACEMENT in B.

5. Antonyms. The term CONTRACTION may be considered equivalent to EXPANSION in one vocabulary or to EXPANSION/CONTRACTION in another.

6. Semantic factoring. This is the most complex situation. The term THERMOMETER in a thesaurus A can only be translated into three uniterms, TEMPERATURE, MEASUREMENT and INSTRUMENT, in thesaurus B.

Smith/[451] has also addressed the problems of human mapping from one vocabulary to another, and has reduced the process to a series of rules. The problems of thesaurus reconciliation in the social sciences are dealt with by Sager et al/[425].

Glushkov et al/[178] distinguish between semantic compatibility and structural compatibility of index languages. Semantic compatibility would imply the ability to represent the same body of knowledge. Two vocabularies in the field of irrigation can be expected to have a high level of semantic compatibility. Structural compatibility refers to the similarity of the vocabularies according to such internal factors as types of terms they use. Two vocabularies that use subject headings followed by subheadings would show some level of structural compatibility. Clearly, there is no direct relationship between semantic compatibility and structural compatibility. A pair of languages may have structural compatibility but virtually no semantic overlap.

The same authors go on to divide the two forms of compatibility further. Semantic compatibility may be reduced to lexical, paradigmatic and syntagmatic compatibility: compatibility in the representation of objects, in hierarchical relations recognized, and in nonhierarchical relations recognized. Structural compatibility can be reduced to morphological compatibility (similarity in the structure of terms) and syntactic compatibility (similarity with respect to the structure of groups of terms).

The convertibility of one vocabulary to another has been a subject studied for many years. Some investigators have looked at the convertibility of bibliographic classification schemes (see, for example, Henshaw/[213], Mishin/[336] and Dahlberg/[102]). In the Soviet Union, the compatibility of the subject indexing used by the All-Union Book Chamber has been compared with that used by five different libraries (Protapova/[398]) and with one large public library (Sobolevskaya/[452]). Ranganathan and Neelameghan/[402] prepared a preliminary study of the compatibility of the Colon Classification and the UDC. The possibility of achieving some measure of compatibility among existing thesauri in the Soviet Union is explored by Antopol'skii and Rudykina/[13]. Others, in constructing a vocabulary, have deliberately made it compatible with one or more existing vocabularies. For example, Jachowicz/[249], describing the construction of a Polish Thesaurus of Hospital Science, pointed out the need to make it compatible with the German Thesaurus Krankenhauswesen as well as with several other thesauri existing in Poland. In other cases, a thesaurus may be made compatible with general terminological standards. Höpker/[222], for example, describes the construction of a German medical thesaurus compatible with the Klinischer Diagnosenschlüssel (KDS), the International Classification of Diseases, and the Systematized Nomenclature of Pathology (SNOP).

In the early 1960's, several related investigations sought to determine how easy it would be to convert the indexing language used by one agency of the United States government to that used by another agency. If the vocabularies of several agencies could be harmonized (e.g., by creating a master table of equivalents among them), the processing of reports produced by these agencies could be integrated and combined subject bibliographies could be generated. This work was discussed in reports by Hammond/[197], by Hammond and Rosenborg/[198, 199], by Painter/[379], by Jaster/[253], and by Datatrol/[104].

The results achieved in these related studies are subject to different interpretations. In general, the Datatrol investigators (Hammond) were optimistic and claimed that their experiments had revealed that useful tables of equivalents among indexing vocabularies could be prepared. Painter was more cautious. She maintained that, while many of the terms in a relatively specialized vocabulary could be mapped to the vocabulary of a larger and more diverse agency, the reverse was not true: only a small percentage of the terms in the diverse vocabulary had equivalents in the smaller. Painter concluded that the specialized vocabulary could usefully be mapped to the more general vocabulary, but not vice versa. It is important to note that this conclusion may have little to do with the size of the vocabulary (in numbers of terms). In point of fact, the specialized vocabulary involved here included over 13,000 terms while the more diverse vocabulary (covering virtually all of science and technology) included about 7000.

Painter also pointed to the practical problems of performing a useful conversion; in particular, the effect of indexing error or lack of consistency on the convertibility of indexing records of one agency into those of a second agency. She claimed that the variable quality of indexing might greatly reduce the utility of any such conversion. For example, term x in vocabulary A may be judged equivalent to term y in vocabulary B. If x is only assigned to 72% of the documents to which it should be assigned, however, many records that should receive the term y (in a conversion from system A to system B) will not receive it. Likewise, errors in B's indexing would cause the same imperfections in a conversion from B to A. If a combined bibliography were generated, the entries appearing under x/y might be many fewer than they should be.

The human intellectual task of converting one vocabulary to another can be both time-consuming and tiresome. However, given that both vocabularies exist in machine-readable form, the possibility exists that some, at least, of the conversion can be done automatically. Wall and Barnes/522, for example, produced a mapping algorithm that would identify exact matches, some variant spellings and some variations in word forms; also, under certain conditions, it would effect mapping on the basis of the cross-references and hierarchical structure of the vocabularies. Some examples of the types of automatic mapping possible are:

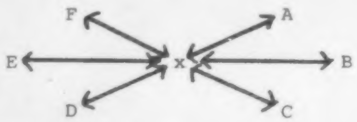
1. Exact match	CALORIMETRY to CALORIMETRY
2. Variant spellings	ELECTRO OSMOSIS to ELECTROOSMOSIS HAEMOGLOBIN to HEMOGLOBIN
3. Word forms	ELECTRIC MOTORS to ELECTRICAL MOTORS
4. Inversions	DISEASES, HUMAN to HUMAN DISEASES
5. Via cross-references	ABATTOIRS to SLAUGHTERHOUSES (where one vocabulary contains a reference ABATTOIRS use SLAUGHTERHOUSES)
6. Via hierarchy	CAROTID SINUS to ARTERIES (where the hierarchy in the mapped-from vocabulary indicates that CAROTID SINUS is immediately subordinate to ARTERIES)

The algorithm will not do all of the conversion automatically but it can greatly reduce the number of intellectual decisions that must be made. As might be expected, Wall and Barnes found that the more alike (compatible) two vocabularies were in structure, the more conversion could be performed automatically. They could automatically map 76% of a sample of Medical Subject Headings, the vocabulary of the National Library of Medicine, to the Agricultural/Biological Vocabulary of the National Agricultural Library because both are structured vocabularies, somewhat similar in format, and with a surprisingly high level of subject matter overlap. In contrast, only 11% was achieved in converting an uncontrolled vocabulary to the Agricultural/Biological Vocabulary and even less success in mapping Library of Congress subject headings to this vocabulary. It should be noted that they did not actually implement the algorithm by computer but performed the entire study through a simulation.

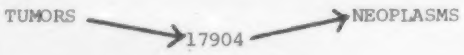
It is important to recognize that, the more vocabularies involved, the more conversion may be possible automatically. To take a simple example, if vocabulary A includes the term x, vocabulary B the reference y use x, and vocabulary C the term y (but not x), it is reasonable to infer than term x in vocabulary A is equivalent to term y in vocabulary C. No such inference could be made automatically if vocabulary B had not been processed.

3.2 The intermediate lexicon

As mentioned earlier, the intermediate lexicon (Coates/87) is nothing more than a switching language. In fact, within a network of information centres, such a device is conceptually similar to an automatic switchboard within a telecommunications network. The situation can be represented as follows



where A-F are information centres and x is the switching language. If the network works properly (e.g., in an online environment), each centre should be able to search the data base of any other using its own vocabulary. A user, then, could sit at a terminal in A and enter terms from A's vocabulary. The terms would be converted to the codes of x and switched to the appropriate terms in, say, D's vocabulary. To use a concrete example, A's user enters TUMORS, which is converted to the intermediate lexicon (x) code, 17904, and switched from this to NEOPLASMS, the equivalent term in D, which situation looks as follows:



Clearly, for this to happen, the vocabulary of each and every centre must be mapped to the switching language. It is also obvious that a reasonable level of commonality in subject matter must exist among the centres, otherwise there would be little point in the activity and little chance of convertibility from one vocabulary to another.

In the example given, TUMORS to NEOPLASMS, the two terms can be considered reasonably synonymous. Some of the transformations will involve identity (JET ENGINES to JET ENGINES), some near-identity (MOUSE to MICE, TIRES TO TYRES, ELECTRIC UTILITIES to ELECTRICAL UTILITIES) and some specific-to-general conversions (STREPTOMYCIN to ANTIBIOTICS). Some, of course, may be extremely general. For example, FOURIER SERIES may map to MATHEMATICAL ANALYSIS.

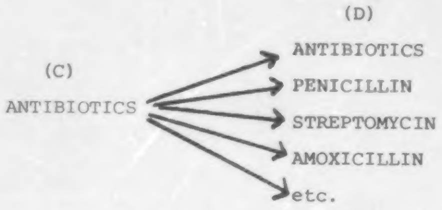
The major problems involved in such mapping are:

1) complete absence of a category of terms in one vocabulary (one vocabulary has specific mathematical analysis terms, such as FOURIER SERIES, while another has no mathematics terms whatsoever);

2) the general to specific situation (one can map FOURIER SERIES to MATHEMATICAL ANALYSIS but how should the reverse be handled: if vocabulary B has only the general MATHEMATICAL ANALYSIS and A has many specific mathematics terms, how does one map from B to A?);

3) vast differences in degree of pre-coordination among vocabularies (one uses mostly single word terms, a second uses more complex, multiword terms).

The third problem, while difficult, is tractable. For example, JET ENGINE NOISE in A may need to be mapped to JETS and ENGINES and NOISE in vocabulary B. The second problem may also be solved by a one-to-many mapping. That is, ANTIBIOTICS may be mapped to a whole group of terms in a second vocabulary:



In this case, entry of the term ANTIBIOTICS in system C will cause the entire group of antibiotics terms in D to be introduced, in an or relationship, into the strategy. The first problem, however, can only be handled satisfactorily by tampering with one of the vocabularies (in the example given, by introducing at least the term MATHEMATICS or MATHEMATICAL ANALYSIS to the vocabulary in which this concept is presently lacking).

The intermediate lexicon can be used in a multilingual environment. Indeed, this is potentially one of its most useful applications. An English thesaurus may map more readily to a French thesaurus, similar in subject matter, structure, level of specificity and degree of pre-coordination, than it will convert to another English vocabulary having quite dissimilar structural characteristics.

Preliminary work on the intermediate lexicon concept was conducted in France (Gardin/[169]). Since then, it has been more fully tested in England, in the field of library and information science, as reported by Horsnell/[228, 229] and Horsnell and Merrett/[230]. These investigators claim that "retrieval performance from indexes constructed from switched indexing is comparable to the retrieval performance from indexes constructed from normal indexing". The implication of this claim is that, given a set of documents indexed by centre A using its own vocabulary, and given a set of requests appropriate to the scope of this data base, it matters little (in terms of such performance measures as recall and precision ratios) whether the data base is searched using vocabulary A or by vocabulary D switched, through the lexicon, to vocabulary A. Readers of this report must draw their own conclusions from the results of this research (Horsnell and Merrett/[230]) as to whether this point is proven definitively or not.

The work of Neville/366, referred to earlier, was also concerned with developing a concordance between several vocabularies, in the same general subject area, and a "supra-thesaurus" consisting of code numbers to represent all of the concepts appearing in the various vocabularies. The procedures used for reconciliation may involve the extensive addition of cross-references to each thesaurus. Some additional terms may also need to be added, but these are kept to a minimum. No existing index terms, in any vocabulary, are deleted.

Assuming the need to reconcile four vocabularies, A, B, C and D, we can begin by taking A as the "source thesaurus". Each term in A is given a unique code and equivalent terms in B, C and D are located and assigned the same code. Thesaurus B then becomes the source thesaurus and new terms (not already encountered in the processing of A) are given codes and reconciled with the other thesauri. The remaining vocabularies, C and D are handled similarly. Presumably one takes the largest thesaurus as source first. As one moves to the smaller thesauri successively, the amount of new reconciliation needed will be progressively reduced.

Beling and Wersig/34 have discussed a wider notion of an "intermediary language" to interconnect a multitude of information systems within the Federal Republic of Germany. Included in their conceptual model is a Bundes-Thesaurus-system, which has a number of subsystems designed to achieve compatibility without imposing undesirable restrictions on the operations of the component information centres. In justifying their approach, the authors claim that

> the time of the large universal systems has gone and the time of more complex intermediary systems mediating between a multiplicity of individual need-oriented systems is approaching. (Page 120)

Somewhat related to the intermediate lexicon are the procedures used within the TITUS II system of the Institut Textile de France (Ducrot/129), which operates in German, English, French and Spanish. A key element is the ability to automatically translate sentences in abstracts from one language to another. This is possible because:

1. A multilingual textile vocabulary was prepared.
2. Abstractors are required to write using a pre-established set of 18 sentence structures ("skeleton sentences").

The procedures resemble the intermediate lexicon concept in that sentences (in whatever language) input to the computer are first translated into a so-called "swivel language" and from this neutral language into any target language.

The subject of multilingual processing is discussed in more detail in 3.7. The long-term objective of automatic translation applied to information services is to develop an approach that is simple and inexpensive enough to permit translation on-demand at time of system output. For example, a Spanish user of a data base will receive abstracts in Spanish (whatever the language they were originally written in) but the translation is performed in "real time" when the user's search is completed.

Diener and Tsuffis/115 have discussed some of the factors to be considered in deciding whether or not to establish an intermediate lexicon among a group of agencies. In addition, they describe a method for testing the feasibility of an intermediate lexicon before extensive resources are committed to its development. By examining records representing documents indexed by several of the agencies, similarity of terminology can be established and problems likely to occur in the reconciliation of the vocabularies can be identified. In essence, the method they recommend entails the establishment of a limited lexicon de facto (based on documents common to several data bases). This, of course, takes indexing policies and protocols into account, and is not quite the same as the reconciliation of vocabularies qua vocabularies, divorced from concern as to how these tools are applied in practice.

3.3 Integrated vocabulary approach

The intermediate lexicon is still a theoretical approach in the sense that no such tool has yet been fully implemented. In an online search environment an alternative is to provide the user with a composite or integrated vocabulary, the terms drawn from all of the data bases processed by a particular information facility. The purpose is a little different from that of the intermediate lexicon in that equivalencies among vocabularies (synonymy or near-synonymy) are not established. Nevertheless, the approach has certain advantages, not the least of which is the guidance it may give on which data base is most likely to be responsive to a particular information need. The integrated vocabulary has already been implemented by the major online service centres in the United States.

DIALINDEX (DIALOG Information Services) and CROSS (Bibliographic Retrieval Services) are files of all the searchable terms appearing in the many data bases of bibliographic records accessible through these services. In response to a term entered by a user online, such a tool will show (a) which data bases the term occurs in, and (b) how frequently it occurs in each. Thus, a user can determine which data base seems to include the most information on a particular topic. DIALINDEX and CROSS will allow searching on word stems and also permit Boolean combinations of terms to be entered. For example, it is possible to enter a combination such as IRRIGATION and BRACKISH WATER and get a count, for each data base, of how frequently these terms appear together.

The System Development Corporation's Database Index is similar in purpose to DIALINDEX and CROSS, as is the Questindex feature of the European Space Agency's IRS service.

One approach to vocabulary merging that does seek term equivalencies has been mentioned by Butler and Brandhorst/66. The Women's Educational Equity Communications Network (WEECN) extracts from many machine-readable data bases those bibliographic references that deal with equal opportunity for women in education. Thus a tertiary data base on this subject is compiled. A "vocabulary guide"/376, which is essentially an alphabetical display, in a single sequence, of all terms (related to WEECN's interests) that appear in each of five data bases, was prepared. For each term, synonyms, near-synonyms and otherwise related terms (from all vocabularies) are displayed, as in the following example:

Fear (MeSH)

RT Conflict (ERIC)
Conflict (PA)
Conflict (SSIE)
Conflict (TEST)
Fear (ERIC)
Fear (SSIE)
Fear (TEST)
Psychological Patterns (ERIC)
Stress (Psychology) (TEST)
Stress-Behavioral Aspects (SSIE)

This shows that the term FEAR in Medical Subject Headings (MeSH) appears explicitly in three of the other vocabularies, is "related to" the term CONFLICT (in four of the vocabularies), to PSYCHOLOGICAL PATTERNS in one and to variants on STRESS in two others. A vocabulary guide of this kind, although loosely structured, has a number of possible uses. It can facilitate the searching of the original data bases, as well as the derived (tertiary) data base, and could be used to aid the conversion of terms from any of the five data bases to the preferred vocabulary, that of the Educational Resources Information Center (ERIC).

Work on the construction of a more sophisticated type of integrated vocabulary has been performed at the Battelle Columbus Laboratories (Niehoff and Niehoff et al[368-371]). The original objective was to create a conversion guide or "synonym table" that would permit online switching from one energy-related data base to another. Experiments were conducted by extracting energy-related terms from 10 data bases and establishing equivalencies among these terms. As in the work described by Wall and Barnes[522], it was discovered that a significant amount of the reconciliation could be performed automatically when the terms were manipulated by computer. The conversion guide constructed in this way could be used to generate a display through which the online searcher can identify appropriate terms. A user could, for example, enter the term LIME and this would generate a display looking somewhat as follows:

Data bases	A	B	C	D	E	F
Terms	LIME	LIME		CALCIUM OXIDES	LIME	CALCIUM OXIDES

which indicates that LIME is a searchable term in A, B, E, that the equivalent term is CALCIUM OXIDES in D and F, and that the concept seems not to appear in data base C. The same display would be generated if CALCIUM OXIDES were entered by the user. As an alternative application, however, the equivalents could be automatically and "transparently" substituted. This means, for example, that the term LIME would be used to search A, B and E, but that CALCIUM OXIDES would be substituted, by computer look-up, for the interrogation of D and F.

The work performed at Battelle permits a searcher to choose any of several switching strategies representing different levels of matching between a term entered and its equivalent in a particular data base as determined by the conversion table: exact match, singular/plural match, stem match (e.g., ELECTRIC POWER and ELECTRICAL POWER), synonym match, broader term match, narrower term match, and so on. In the latest version of the programs the display generated includes a record of the type of match that led to the terms shown for each data base. Presumably, to obtain high precision of search results, one would select a fairly exact matching strategy. To improve recall, however, one would relax the matching requirements.

In later work at Battelle[369, 370], the scope of the project was expanded to other subject areas, encompassing 15 vocabularies in science and technology, the social sciences and business. An online integrated vocabulary in the range of 40-50 million characters has been built and tested in an experimental setting. The investigators claim that the approach used is as relevant in a multilingual setting as it is in a monolingual one (one of the vocabularies worked with is an English/Spanish/Portuguese vocabulary in ferrous metals). They have also discovered varying degrees of vocabulary compatibility in different subject areas, business vocabularies being more compatible than those in the sciences. The approach adopted by Battelle has yet to be implemented in a real information service setting. Testing of the approach was underway in 1983.

The integrated vocabulary is a little different from the intermediate lexicon approach because no "neutral" switching language has been created. Within an online network, however, since vocabulary equivalencies are established in both procedures, the end results of the two approaches may not be too much different. Dahlberg/[103] has referred to a similar method as a "compatibility matrix".

Somewhat related to the work of Niehoff are the investigations reported by Williams and Preece/[546] on an automatic "data base selector". By storing a merged vocabulary, drawn from many data bases, and developing an appropriate algorithm, such a device can be used to "test out" a rough search strategy before a particular data base is selected for interrogation. Based on the frequency with which terms occur in these data bases, the selector can point to the data base, or data bases, offering the best possibility of satisfying a particular information need.

3.4 Microthesauri

The term microthesaurus has no universally accepted definition. It is frequently used loosely to refer to a rather small and specialized thesaurus. In its original application, however, a microthesaurus was a specialized subset of terms extracted from, and therefore compatible with, a larger thesaurus, and this restricted usage is certainly to be preferred. A microthesaurus, then, can be defined as a specialized vocabulary that maps to, and is entirely included within the hierarchical structure of, some broader thesaurus.

Consider a network of specialized information centres in, say, agriculture. The hub of the network is a co-ordinating centre maintaining a general thesaurus and a general data base in the agricultural sciences. Each centre operates in a specialized subject field: irrigation, dairy science, cacao, rice culture, and so on. Each produces its own specialized data base and printed bibliography. The co-ordinating unit produces a general index in agriculture.

The problem is that the satellite centres need to index their specialized literatures with much greater precision than is true of the co-ordinating centre. In the general data base, some of the terms contributed by the specialized centres should appear at a higher level of specificity. The co-ordinating centre needs a complete but somewhat general vocabulary covering all of agriculture. Each satellite needs a highly specific vocabulary for its field of specialization but also needs to draw on less specific terms from other areas of agriculture that impinge upon its own specialized area.

This is an ideal situation for the use of microthesauri. Each satellite develops a highly specific vocabulary for its specialty and this maps to the general thesaurus, fitting entirely within the hierarchical structure of that thesaurus. In indexing its collection, the satellite uses its microthesaurus and draws more general agricultural terms from the parent thesaurus. The indexing of the satellites is completely compatible with the needs of the general data base: machine-readable records from each can be merged by the co-ordinating unit, specialized terms being automatically mapped to the appropriate more general terms (e.g., CHAROLAIS might be converted to BEEF CATTLE). In this situation, the vocabularies of all specialized centres are compatible, compatibility being achieved through the structure of the parent thesaurus, but mapping is only one way (from the specific vocabularies to the general, not vice versa).

An example of the integration of two thesauri, one specialized and one general, so that one becomes a "satellite thesaurus" of the other, can be found in Hammond/[196]. The integration of a "base" thesaurus in chemistry with supplementary special thesauri is discussed by Bauer/[31]. Neicu/[363] has referred to the need to make satellite thesauri (oriented to specific subject areas) compatible with the Romanian national thesaurus, and to develop standards to promote this compatibility.

Soergel/454 pointed to the need for a "core classification" from which "extended schemes" could be developed for specific applications. He later expanded this into the notion of a Universal Source Thesaurus, from which others can be derived/455. Very recently the British Standards Institution (BSI) has produced something resembling such a tool; its ROOT Thesaurus/60 (see also Dextre and Clarke/114) is intended as a comprehensive indexing and searching tool in the field of technology. ROOT contains some 11,800 descriptors plus an additional 5500 entry terms. In fact, it is more a thesaurofacet, combining the features of a faceted classification scheme with those of a thesaurus, than a conventional thesaurus. ROOT is available in both printed and machine-readable form. Because it has the BSI behind it, it is reasonable to suppose that it will be kept current. The reason for its mention here is that the compilers claim that the tool can form a base for specialized thesauri developed by other agencies. To quote from a BSI brochure on ROOT:

> The combination of a logical structure and computer generation system means specialized thesauri can be produced quickly and easily, to suit your particular requirements. All you have to do is select and expand the desired schedules, take extracts from other supporting areas and discard any irrelevant sections. The computer will then do the rest to provide you with a custom-made thesaurus at minimum cost.

Several microthesauri developed from ROOT would be compatible with each other as well as with the parent tool.

While not microthesauri as such, a number of specialized vocabularies have been built to complement a more general controlled vocabulary. The SfB system, in the field of building construction, is an example of a specialized classification scheme designed to complement and be compatible with the UDC (Giertz/176).

Sokolov et al/457 have described and tested procedures for the formulation of a group of interrelated thesauri fitting within the conceptual structure of a "universal paradigm". The faceted approach adopted recognizes three levels of subject access: a universal paradigm of facets common to all thesauri whatever the subject, type paradigms, which extend the universal paradigm into broad subject areas, such as transportation, and the specific paradigms, which extend the type paradigms to a greater level of detail, e.g., navigation as a subset of transportation. If this were carried to its logical conclusion, one could presumably create a cohesive structure of interlocking thesauri, at various levels of specificity, suitable for application in a multifaceted national information programme. Various levels of microthesauri would exist within the complete structure. Sokolov et al describe the construction of two mutually compatible thesauri at the lowest level of the structure, one on navigation and one on sea ports. The work was performed at the Leningrad N.K. Krupskaya State Cultural Institute. On a more theoretical level, Vitukhnovskaya/517 has discussed principles involved in constructing a whole system of compatible thesauri.

The ROOT thesaurus, previously described, could conceivably serve as the superstructure of a national information system, the various specialized vocabularies being developed to fit within this superstructure. The entire system might comprise three, four or five levels, as in the following example:

Multidisciplinary superstructure
Science and technology
Energy
Renewable energy sources
Solar energy

Thesauri would exist at each level. Thus, the solar energy thesaurus (suitable for use in a solar energy research institute) would be a microthesaurus of the renewable energy thesaurus, which would be a microthesaurus of the energy thesaurus, and so on. Since the entire structure would be interlocking, terms from higher generic levels can readily be selected by any specialized centre (e.g., the solar energy institute may need to draw some general terms from, say, social science and medicine).

Speaking of the situation in East Germany, Ellmann/[140] refers to the need for constructing specialized thesauri from existing ones, and Plante et al/[391] have described a French macrothesaurus from which various subject-related editions can be derived. The *Macrothesaurus des Sciences et Techniques* covers mathematics, the physical sciences, biomedicine, the earth and space sciences, industry, agriculture and some aspects of economics, law and geography.

Unfortunately, most existing index languages have been developed completely independently. While they may conform to common guidelines in their structure, they have not been integrated or interconnected within a common structure. Of course, the polythematic thesauri needed to provide the superstructure have not previously been available. At least, thesauri that might have performed such a function (e.g., the monumental *Thesaurus of Engineering and Scientific Terms* in the United States) were not developed in a form suitable for such application and have not had much influence in this particular way.

Referring to the current situation in the Soviet Union, Shemakin and Kulik/[440] have pointed to the inefficiencies associated with the continued proliferation of specialized thesauri that are not interlocked within a single superstructure:

> The problem of creating this fundamental thesaurus and general compatibility principles is becoming more acute in that at present various branch thesauruses are being more and more actively developed. If these are not placed under some sort of supervision--and the means of this supervision should be the polytechnical thesaurus--then when the various ministerial IRS develop their own autonomous thesauruses and index their document collections using them, organizational and technical difficulties will arise as the result of the necessity of adjusting these available thesauruses to the normative terminology and the necessity of reindexing the existing document files. (Page 40)

This section has described the development of specialized vocabularies to fit within an existing superstructure. The reverse procedure, building a superstructure to encompass existing vocabularies, is dealt with in the next section.

3.5 The macrovocabulary

The macrothesaurus approach (see Wolff-Terroine/[557]) is conceptually similar to that of the microthesaurus but the method of implementation is virtually reversed. The idea is simply to create a kind of generic superstructure of terms that will subsume a group of thesauri, or other types of vocabulary, in diverse subject fields. In fact, when this approach was first proposed, as a means of interconnecting several vocabularies used by government agencies in the United States, it was referred to as a "common subsumption scheme". The subsumption scheme was, in fact, developed under the auspices of the Committee on Scientific and Technical Information (*COSATI Subject Category List*/[501]) and published in 1964. The list is used to arrange the *Government Reports Announcements* of the U.S. National Technical Information Service, the major bibliography of technical reports within the United States. It has been used to "partition" the terms in several thesauri (e.g., the *Thesaurus of Engineering and Scientific Terms*) into broad subject categories.

The Broad System of Ordering or BSO/[61], developed as an element of the UNISIST programme, is the most obvious manifestation of the macrovocabulary approach. As reported by Toman and Lloyd/[486], the BSO was conceived of as a tool to allow the interconnection of information systems, services and centres using diverse index languages, a tool to permit the consistent representation of subject fields and subfields at a generic level, and a tool to categorize information centres and other sources of information, thus permitting the establishment of a world register or referral centre. Essentially, then, the BSO was intended primarily as a switching language (but merely as a general superstructure rather than the more specific switching approach embodied in the intermediate lexicon), although other potential applications were also recognized, including the broad organization of multidisciplinary publications, data bases and, possibly, libraries.

The BSO exists in the form of a classification scheme, with notation, having just over 4000 terms. Justification for including a term was "institutional warrant" (Coates/[85]). That is, a term was included if it was known to represent at least one "organized information source" (e.g., information service or centre, abstracting or indexing service, university department).

The merits and demerits of the Broad System of Ordering have been widely discussed in the literature (Wellisch/[530], Foskett/[161], Vickery and McIlwaine/[515], Perreault/[386], Soergel/[453], USSR State Committee/[503], Dahlberg/[100], Coates/[86]). Many of the criticisms relate to relatively minor structural features, more relevant to the ordering of physical artifacts (e.g., books on library shelves) than to the indexing of machine-readable records. More important is an understanding of the limitations of the scheme in terms of its possible applications.

The BSO does have potential uses. It could be used to organize a relatively small multidisciplinary library or to segment a multidisciplinary data base (such as the French PASCAL). It would certainly be useful for the categorization of information sources within a general referral centre (although not specific enough to support the activities of a specialized referral system such as INFOTERRA), and perhaps for categorizing the contents of data bases within a large online network. BSO could also be used to switch broad categories of records between machine-readable data bases (e.g., "metals industry" records appearing in a biomedical data base could be transmitted to an information centre serving this industry) but only if each data base tagged its records with BSO codes, as well as its own index terms, or its own vocabulary were mapped to BSO (either requirement being relatively expensive to implement). Because of its comparative lack of specificity, the BSO is unsuitable as a means of converting the detailed indexing of one agency into the detailed indexing of a second. For the same reason, it would not serve the network purpose of allowing a vocabulary-specific search strategy to be converted into the vocabulary of another data base. The BSO, then, cannot perform the function of an intermediate lexicon as defined earlier. (Vilenskaya/[516] distinguishes between intermediary languages and switching languages. She would regard BSO as a switching language but not an intermediary language.)

There are other potential uses, however, that appear not to have been explored. Suppose, for example, that all journal publishers could be prevailed upon to assign BSO codes to all articles published, and that these codes were carried in a general machine-readable file (such as the three major data bases of the Institute for Scientific Information, covering the sciences, social sciences and humanities). The general file could then be used to "feed" the more specialized indexing/abstracting services (thus obviating the need to define the scope of these services almost exclusively in terms of a limited set of journal titles and facilitating access to the more fugitive literature appearing in unlikely sources) and even to "feed" specialized information centres on an interest profile basis.

Another possibility, given that records in various data bases include BSO codes, as well as their own index terms, is to use the superstructure to suggest useful search terms to retrieve information on particular subjects from a variety of data bases. Some research on this approach was conducted almost twenty years ago in an effort to achieve some compatibility or convertibility in the handling of technical reports within the United States. The idea is simply to develop, for each of the broad terms from the superstructure, a profile of the terms from a specific data base that are associated most frequently with the broad code. For example, Table 3, taken from Hammond/197, shows two term profiles for the subject category Heat Transfer, one taken from the data base of the National Aeronautics and Space Administration (NASA) and one from that of the Defense Documentation Center (DDC). Such a display, for example, could be used within an online network to bring to the attention of a DDC user the NASA terms relevant to a search on heat transfer (or vice versa). While experiments along these lines have been conducted, as far as the authors are aware, this approach has never been implemented in an operational setting.

Whether any of these uses actually materialize is largely dependent on Unesco's ability to conceive and promote innovative applications of the BSO and its willingness to commit resources to keep the system up to date. At the time of preparing this report, no actual applications of the scheme have been identified.

In the development of its National Technical Information System, the Soviet Union seems to be favoring what is essentially a macrothesaurus approach. Pershikov/387 refers to the need to make interfacing subsystems compatible with the parent system, at the vocabulary level as well as in hardware, formats, computer programs and technical procedures. The interfacing subsystems are automated scientific and technical information systems of all-union, republican, regional, and specialized information agencies. The situation may be handled by means of an information retrieval language having two levels: a classificational level in the form of a subject heading authority for the topical areas, and a descriptor level in the form of a set of descriptor languages corresponding to the various topical areas. The subject heading authority is used to organize resource files and to distribute requests among the topical subdivisions (i.e., act as a macrothesaurus) while the descriptor languages are used for indexing and retrieval of documents in a precise subject area.

The need is further elaborated on by Chernyi/76:

> We believe that a single information system of subject headings for the National Information System must be a collection or composite of the systems of subject headings of VINITI (for published scientific documents), the AUSTIC* System (for unpublished scientific documents), and the

---

* AUSTIC is the All-Union Scientific and Technical Information Centre. Its subject headings are used for categorizing unpublished reports on research and development projects carried out in the Soviet Union.

TERM PROFILES
DDC

1935			HEAT TRANSFER
264	92	702	Ablation
1578	119	511	Aerodynamic Characteristics
519	57	502	Aerodynamic Configurations
525	226	817	Aerodynamic Heating
722	77	528	Air
541	103	640	Atmosphere Entry
291	79	658	Blunt Bodies
331	54	554	Bodies of Revolution
49	26	620	Boiling
783	209	754	Boundary Layer
1007	90	514	Combustion
253	51	576	Compressible Flow
360	60	568	Conical Bodies
215	119	780	Convection
21	13	563	Cook-off
128	64	706	Coolants
591	196	773	Cooling
900	115	597	Cylindrical Bodies
196	56	629	Enthalpy
10	9	563	Film Boiling
27	15	567	Film Cooling
20	10	511	Flat Plate Models
975	142	637	Fluid Flow
476	79	595	Fluid Mechanics
208	34	506	Fluids
444	66	562	Friction
1208	244	736	Gas Flow
1673	178	613	Gases
366	55	544	Heat
226	118	773	Heat Exchangers
502	66	544	Heating
62	17	500	Hemispherical Shells
1820	132	514	High Temperature Research
451	109	676	Hypersonic Characteristics
437	126	712	Hypersonic Flow
217	43	556	Hypersonic Wind Tunnels
300	68	620	Hypervelocity Vehicles
429	169	778	Laminar Boundary Layer
26	13	540	Liquid Cooled
169	47	607	Liquid Metals
514	65	537	Liquids
168	31	513	Mach Number
7383	369	534	Mathematical Analysis
167	39	567	Nose Cones
25	18	615	Nucleate Boiling
214	43	558	Pipes
3041	222	570	Pressure
31	15	552	Radiators
28	12	514	Reactor Coolants
523	85	600	Reentry Vehicles
179	38	552	Reynolds Number
414	61	552	Rocket Motor Nozzles
950	83	502	Rocket Motors
428	53	513	Shock Tubes

NASA

1100			HEAT TRANSFER
237	56	550	Ablation
175	58	598	Aerodynamic*Heating
109	53	635	Boiling
666	201	714	Boundary*Layer
261	52	518	Conduction
252	120	716	Convection
450	109	622	Cooling*/Noun/
198	53	561	Enthalpy
304	61	535	Flatness, *Flat
2537	350	659	Flow*/Noun/
615	86	513	Fluid*/Noun/
19	12	509	Free*Convection
1811	170	505	Gas*/Noun/
1214	[illegible]	754	Heat*/Noun/
98	41	591	Heat*Flux
127	116	786	Heat*Test
481	99	589	Heating, *Heated
691	38	619	Hypersonics
392	127	675	Laminar
626	[illegible]	507	Layer
86	42	612	Mass*Transfer
799	[illegible]	503	Nozzle*/Noun/
22	[illegible]	570	Nucleate
23	[illegible]	[illegible]	Nusselt*Number
651	[illegible]	513	Plate
594	83	509	Point*/Noun/
64	34	[illegible]	Prandtl*Number
43	26	587	Radiative
267	67	576	Reynolds*Number
240	48	510	Skin
295	107	672	Stagnation
2834	261	547	Temperature*/Noun/
140	62	640	Temperature*Distribution
34	17	520	Temperature*Profile
1255	154	550	Thermal*/See*Also*Thermo-, *Heat/
187	40	501	Thermocouple
677	309	808	Transfer*/Noun/
362	82	583	Turbulent
448	70	511	Viscosity
419	82	562	Wall*/Noun/
50	34	628	Wall*Temperature

Table 3

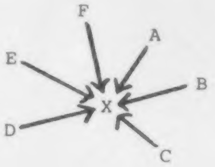
"Term profiles" for the broad category Heat Transfer in two vocabularies. The 1935 and 1100 refer to the frequency with which the "parent term" has been used in indexing. Figures in the three columns represent (a) how frequently the specific term has been used in indexing, (b) how frequently the specific term has been used with the parent term, and (c) an "association factor" reflecting the frequency of co-occurrence of specific term and parent term. Reproduced from Hammond/[197] by permission of the National Bureau of Standards.

> system of subject headings of all specialized information systems. For convenience all these systems should be equipped with an alphabetical subject index to link them as well as a UDC index. (Page 7)

In France, the macrothesaurus approach was once seen as the solution to the linking of various specialized information centres and to achieving some measure of compatibility among the vocabularies of these centres. As reported by Michel/332, the approach has been abandoned. It was found not to work because of the problems of "polyhierarchy" (i.e., topics falling legitimately into various hierarchies).

3.6 Algorithmic approaches

The methods so far discussed relate primarily to the mapping of one vocabulary to another. Somewhat related work has been performed at the data base level. The situation arises in the construction of a tertiary data base--one derived by drawing records from other data bases; it can be illustrated as follows:



Suppose, for example, that X is a tertiary data base on solar energy, drawing its bibliographic records from the existing data bases A-F, covering energy in general, physics, engineering, and so on. Given that A-F exist in machine-readable form, a "profile of interest" for X can be matched against these sources on a regular basis, in much the same way that a user interest profile is matched against data bases in selective dissemination of information (SDI) systems. Thus, A-F "feed" the tertiary data base X on a continuous basis.

In theory, this process could be carried further so that, when records are transferred into X, the appropriate index terms from the X vocabulary are automatically assigned to these records. This could obviously be achieved by developing a complete set of mappings between the vocabularies of A-F and the vocabulary of X. Clearly, this would be an expensive proposition. An alternative is to seek ways of assigning X terms automatically in some other way. This field of research is somewhat removed from compatibility per se. It can be considered a branch of automatic indexing or automatic classification, subjects that have extensive literatures in their own right (see, for example, Stevens/463, Sparck Jones/460 and Fangmeyer/149). Nevertheless, it is important to recognize that research on automatic classification has been performed with the goal of establishing "automatic database translation and mapping procedures." For example, Cahn and Herr/69 have described techniques for automatically classifying documents in a tertiary data base in the energy field. In essence, index terms/text phrases in the records are used to predict which category a particular record should be assigned to. The terms/phrases used as "predictors" of a category are those occurring frequently in records that human indexers have assigned to that category. Thus, the process tries to achieve algorithmically what the human classifier does intellectually. This type of mapping should be regarded as purely experimental at the present time.

Another area of related research has to do with the automatic linking of synonymous or semantically related terms within a particular data base. The main reason for doing this is to suggest possibly useful terms to a user of the data base (e.g., a searcher at an online terminal might enter a particular group of terms and the "system" uses these leads to draw him to other terms or to add further search terms automatically). The pioneering work in this area was Stiles' "associative indexing" (Stiles/467, Salisbury and Stiles/427) but many similar techniques have been developed. The most recent, and in many ways the most sophisticated, is a natural language interface now being developed for use with MEDLARS (see Doszkocs/124 and Doszkocs and Rapp/125).

3.7 Multilingual processing

The technological developments that facilitate co-operation among information services in general make it increasingly possible for such co-operation to take place at an international level. An impediment to effective global information transfer is the multiplicity of national languages in use.

Of course, languages, like empires, tend to rise and fall. Greek and Latin no longer exert much worldwide influence. Earlier in the twentieth century (Hulme/232) German was the major language in the sciences. Today this position is occupied by English (Narin and Carpenter/348). A scientist who reads English with ease is thus in a favoured position vis-a-vis his colleagues without this facility. This advantage is even more pronounced in relation to formal information services: the indexing and abstracting services, especially those in machine-readable form, are mostly produced by the English-speaking countries. Iljon/236 reports that, of 337 data bases in use in Europe in 1976, 76% used English as a carrier language and 99% of the use of all data bases was achieved through the medium of English. The worldwide adoption of English as the lingua franca of scholarly communication, however, is not an acceptable means of removing the "language barrier", for political and nationalistic reasons if for no other.

The ability to handle publications in many languages, and to accommodate several languages in the design of information services, is becoming more important as truly international systems become operational. A most obvious example can be found in the activities of the Commission of the European Communities. Not only are several languages involved, but this multilingualism is mandated by law. Solutions to this problem in the European Community are wide ranging: expansion and improvement of language teaching, research into terminology, development of terminology banks, publication of multilingual glossaries and thesauri, refinement of machine translation procedures, and so on/377.

Language differences may become more of a barrier to effective information transfer in the future than they have been in the past. As Price/397 pointed out some years ago, the pace of progress in science and technology has already begun to slow in the most developed countries. In parts of the developing world, on the other hand, a rapid acceleration is evident in scientific and technological achievement. While many of the researchers from these countries still choose to publish abroad in journals perceived to be more prestigious (and predominantly in English), it seems only a matter of time before the literature of science (and of scholarship in general) becomes scattered over more languages than ever before. Goetschalckx /179, in fact, has suggested that the proliferation of "significant" languages may force us to seek different approaches in information handling, perhaps even the adoption of some artificial language:

> At the theoretical level, it is not unreasonable to consider the usefulness of a universal intermediate language, whether it be a numerical language, esperanto, or something else. When the developing countries take their place, there will be needs out of all proportion to the intellectual and linguistic capital needed to cope with the work of human translation alone. (Page 146)

Multilingual processing in documentation activities can be considered at several levels. Most obviously, we can distinguish between:

1. Pointers to the content of foreign language literature in the form of bibliographic records in printed or machine-readable data bases.

2. Intellectual access to the content of the documents themselves.

Significant improvements at the second level have occurred in the last thirty years. More material is translated than ever before (including some periodicals that are regularly translated cover-to-cover), translation is facilitated by improved glossaries and other aids, and translations centres (national and international) maintain records of translations that have been made, announce these translations, and offer copies for sale. This level of access, however, is peripheral to the subject of compatibility and therefore not of major concern in this report.

Multilingual access to bibliographic records can itself be broken down to at least three levels:

1. Translation of titles

2. Translation of index terms

3. Translation of abstracts.

Several of the indexing and abstracting services include titles in the original language and in translated form. Data bases containing index terms in more than one language are rare indeed. The French PASCAL data base (accessible online as PASCALINE) does have some subfiles in which keywords appear in English as well as French, allowing searches to be performed in either language.

In general, however, a data base indexed by means of terms selected from language A can only be searched by means of terms in language B if:

a) A's terms are humanly translated into B's, or
b) A's terms are automatically translated into B's, possibly through a multilingual glossary or thesaurus.

Bilingual and multilingual controlled vocabularies are not new. Muench/343, for example, described the development of a bilingual (English/Spanish) index to six such vocabularies in the field of biomedicine. The National Library of Medicine's Medical Subject Headings now exists in several languages, allowing the MEDLARS (or MEDLINE) data base to be interrogated in languages other than English.

A notable multilingual thesaurus is that developed by the French Road Research Laboratory, for use by the road research laboratories of OECD member countries, which uses arrowgraphs to display term relationships and a superimposable grid to show term equivalents in various languages (Van Dijk/506). A similar approach was used by the Commission of the European Communities in the construction of a trilingual Metallurgy Thesaurus (Rolling/412, 413). The development of three multilingual thesauri in the field of agriculture, also within the Commission, is described by Iljon/235.

The compilation of multilingual thesauri has been made easier by the appearance of the UNISIST guidelines for multilingual thesauri/497 and the development of computer programs to manipulate such thesauri (Rolling/411). While guidelines reduce the problems, they certainly do not eliminate them. Some of the problems of constructing a multilingual thesaurus are discussed and illustrated by Pigur/388, and Bykova and Muzrukov/67, with special reference to translation to and from Russian.

One result of increasing activity in international information systems has been a resurgence of interest in machine (or machine-aided) translation and remarkable improvements in this technology. Rolling/414 describes ongoing efforts in this area within the Commission of the European Communities. Systran, a system developed by the World Translation Center, is said to have an "intelligibility rate" of 78% (Van Slype/507) and translating speeds of 300,000 words per hour (Toma/485). Rolling implies that this can be pushed to over 90%. He makes reference to a new system, Eurotra, being developed by the Commission itself, which has as its goal an intelligibility rate of 99%. The system will be designed to handle all languages of the Commission countries (King and Perschke/264). Rolling points out that, within an international network such as the Euronet DIANE, it is not necessary to translate for input purposes. Instead, machine translation (e.g., of abstracts) can be achieved at the time of output from the system. The Commission has also developed a software package, Astute, for the compilation, maintenance and printing of multilingual thesauri, as well as a multilingual terminology data bank, Eurodicautom, which is accessible online.

Multilingual processing is not necessarily infeasible in a natural language environment. For example, Salton/428 has reported on the use of his SMART system in the processing of English and German texts. By building an internal "thesaurus" (not a conventional structured tool but one including synonym tables and tables of conceptually related words) in the two languages, it was possible to achieve cross-language processing. German documents could be searched using English language queries in sentence form, and vice versa for German queries. Salton claims that the results were almost as effective as processing with a single language. An approach to constructing a German/English thesaurus automatically, using bilingual and monolingual dictionaries as input, has been described by Bollmann and Konrad/46.

Controlled vocabularies coupled with automatic translation (or, alternatively, the existence of multilingual thesauri) make it increasingly possible to contemplate the production of printed indexes in various language editions. Gibb and Phillips /175, quoting d'Olier, refer to the possibility of producing an abstracts journal in two separate language editions at a cost of about 50% more than a single language edition. Additional relative economies would be possible as further language editions were added.

In a multilingual community, however, it may be more expedient to abandon multilingualism (on economic grounds if no other) in favor of the adoption of a single language for indexing and searching. Thus, Russian is the working language in many of the information systems within the CMEA countries, especially systems having an element of internationalism (Sorokin/459). The same is true for various systems within the United Nations. For example, in AGRIS English has been adopted as the carrier language. This offers the obvious advantage of allowing consistency throughout the network. On the other hand, it also means that the indexers and searchers at each national centre must have a good command of the English language and of the technical terminology of agriculture in particular. Wellisch/535 claims that the ability to read English, or even to speak it, does not in itself guarantee that one can index or search a data base effectively in that language. Moreover, as Wolff-Terroine/558 has pointed out, "International co-operation requires a political and psychological consensus which is difficult to achieve if there is only one working language".

Bilingual or multilingual vocabularies can improve the situation. If a bilingual thesaurus exists in machine-readable form, a national centre could index and search in its own language and still prepare input that is completely compatible with the needs of the parent system. In the case of the International Nuclear Information System (INIS), for example, the system thesaurus exists in French and Russian (Semenov et al[439]) versions, as well as in English.

For international operations, of course, a strong case can be made for the adoption of a "neutral" language. Classification schemes, such as the Universal Decimal Classification (UDC), satisfy this requirement since the class numbers in such a vocabulary are not language-dependent. The UDC has great advantages: it exists in several languages and is used in many countries. Nevertheless, its use within the United States, still the world's largest producer of scientific and technical information, is insignificant.

Lloyd[290, 291], recognizing that existing information services in an international network would be reluctant to abandon vocabularies already in use, once urged that UDC be adopted as a standard switching language (in effect an intermediate lexicon) with which each existing vocabulary could be reconciled. Use of the UDC for this purpose has not found favour, partly because of wide dissatisfaction with the rather uncontrolled procedures by which this tool is updated. In fact, the development of the BSO can be considered a deliberate rejection of the UDC as an international switching device.

Classification schemes can provide, as it were, multilingual pointers to the subject matter of documents. They do not, of course, provide a detailed level of access to the contents of these documents. A different type of neutral language is the "constructed language" such as Esperanto. If adopted worldwide as the language of scientific publishing, primary as well as secondary, there is no doubt that a constructed language could greatly simplify the international exchange of information. There has been some recent evidence of renewed interest in such artificial languages, and Frank et al[163] have presented cogent arguments in favour of their use in international documentation activities. Nevertheless, while a few isolated pockets of enthusiasm do exist, worldwide interest in this approach seems at best lukewarm.

Before leaving the subject of multilingual processing, it is necessary to reemphasize that the intermediate lexicon approach, if it works at all, is equally applicable in a multilingual environment. To reconcile six controlled vocabularies with an intermediate switching language may be only marginally more difficult if these vocabularies involve the use of, say, four national languages than if all six use a single language.

3.8 Other aspects of terminological compatibility

Terminological compatibility in general (unlike compatibility among index languages) is somewhat tangential to the main thrust of the present report. Nevertheless, it seems important to mention the fact that the standardization of terminology, including some degree of international harmonization, is an area of increasing interest and activity.

This type of standardization has traditionally been promoted through the use of technical dictionaries and glossaries. Multilingual versions of these tools assist the work of the translator and further consistency in translation. Such tools, however, are expensive to produce and tend to become obsolete very quickly.

An alternative, and one of growing importance, is the terminology data bank which De Besse/[109] refers to as:

> A kind of living multilingual electronic dictionary containing hundreds of thousands of technical and scientific terms together with the appropriate terminological information. (Page 133)

In the last several years, many such data banks have been set up, mostly by governmental agencies but some within private industry. Several can now be accessed online. Schulz/[436] describes the TEAM system developed to aid translation activities at Siemens AG, Paré/[380] a French/English data bank at the University of Montreal, and Felber/[151] the use of terminology banks and other methods of achieving harmonization of technical terminology within the countries of the Council for Mutual Economic Assistance (CMEA). Perhaps the most important of terminological data banks is Eurodicautom, a multilingual source of terminology operated within the European Community and accessible online (Goetschalckx/[180]). A more complete discussion of this subject can be found in Terminological Data Banks/[479]. In particular, the problems involved in the exchange of terminological data (including international exchange) are detailed in several papers in this volume. Design factors are discussed by Morton/[341].

The single most obvious manifestation of international concern for the standardization of terminology was the establishment in Vienna in 1971 of the International Information Centre for Terminology (INFOTERM), an integral component of the UNISIST programme, affiliated with the Austrian Standards Institute (Felber et al/[154]). INFOTERM's goal is to co-ordinate terminological activities worldwide through:

collecting relevant publications in all languages
preparing bibliographies of terminological publications
maintaining information on libraries having relevant collections
disseminating information on terminological publications
providing advisory services, particularly to institutions in developing countries
organizing meetings
promoting the interconnection of terminology data banks.

Through INFOTERM's initiative, a loosely structured network of centres involved in terminological work, known as TermNet, has existed since 1979 and plans are being made for the reciprocal exchange of data in machine-readable form. According to Riggs/[409] some 16 terminology banks were known to be in existence in 1977.

Beling/[33] has pointed out that the development of a true network of terminological data banks, linked through telecommunications, is impeded by lack of compatibility in the content and data structures used in the various centres. He describes the characteristics of MATER, a magnetic tape interchange format for terminological/lexicographic data, serving the same function for terminology data banks that UNIMARC does for bibliographic data bases. MATER (Magnetic Tape Exchange for Terminological/Lexicographical Records) was drafted by the German Institute for Standardization (DIN); it was later elaborated as an international standard, ISO/DP 6156. Work is now proceeding at INFOTERM on the identification of mandatory and optional data elements for terminological records (Felber/[152, 153]).

The Unesco project, INTERCONCEPT, launched in 1977, has as a major goal the establishment of a bank of terms, with definitions, in the social sciences (Dahlberg/[101], Vasarhelyi/[509]).

A good picture of the present status of work on the standardization of terminology can be found in two recent publications from INFOTERM/480, 481.

3.9 Problems relating to index languages

Several reconciliation problems were identified earlier: wide differences between vocabularies in degree of specificity, differences in degree of pre-coordination, and complete absence of a needed concept in a particular vocabulary. There are other problems associated with the use of computer processing to achieve some level of reconciliation automatically. Most obvious is the fact that in different vocabularies a term may have quite different usage (e.g., STRIKES could relate to a labour dispute in one and to a military operation in a second) or, alternatively, a slightly different nuance. For example, Kirtland et al/267 cite the difference between the term BUDGETING in a general sense, perhaps related to personal finances, and BUDGETING specifically restricted to a formal activity performed by an industrial organization, while Smith/451 cites the difference in scope between ELASTICITY in an engineering thesaurus and the same term in a medical thesaurus.

It seems obvious that, given the desire to reconcile controlled vocabularies, the whole process would be considerably simplified if more organizations followed the same conventions. Two thesauri following the ISO *Guidelines for the Establishment and Development of Monolingual Thesauri* (ISO 2788-1974), or following a compatible national standard (the U.S. standard, ANSI Z39.19-1974, *Guidelines for Thesaurus Structure, Construction and Use* is, in almost all respects, fully compatible with the ISO standard), are much more likely to be readily convertible than two vocabularies constructed on different principles or following different standards. Nevertheless, while the standards do promote consistency in several important respects (e.g., by requiring direct entry rather than inversion, by establishing singular/plural conventions) they fail to grapple successfully with at least two of the most important characteristics of vocabularies: the appropriate level of specificity and the degree of pre-coordination. Variations in these characteristics are among the greatest barriers to the convertibility of vocabularies.

The specificity problem, in fact, is not amenable to solution through standardization since it is unreasonable to expect that any organization would want to develop every term hierarchy to the same level of specificity. The pre-coordination problem is somewhat more tractable. The fact remains, however, that universally acceptable and fully explicit guidelines on this matter have not yet been produced (see, for example, Jones/257).

A further problem relates to updating. To maintain its utility, a vocabulary needs to be kept current. This is not necessarily an easy task. Currency would seem to be especially important in the case of a switching language. But the problems of updating are compounded when input from a whole group of information centres is involved. It is not enough to build a switching language; a mechanism must exist for keeping it up-to-date at all times (Foskett/161).

The problems of vocabulary convertibility are frequently underestimated. While it may be possible to convert vocabulary A to vocabulary B, in practical application this transformation may be fraught with dangers. Unfortunately, it is not enough to consider the vocabularies of system A and system B without also considering the underlying indexing policies and protocols of the two organizations (i.e., how these vocabularies are applied in practice). To take a very simple example, in one system the term OPHTHALMOLOGY may be used very generally to include all aspects of diagnosis and treatment of eye diseases. In a second system, however, the term is used exclusively to represent the profession of ophthalmology. Interconversion of these vocabularies, at the surface level, could produce some very strange results were the indexing records of centre A converted to those of B.

The problem is likely to be especially acute in any automatic mapping operations. Suppose that a mapping algorithm converts FUNGAL DISEASES in A to what appears to be the closest equivalent term in B, namely FUNGI. Assuming that both data bases are biomedical, such a mapping seems reasonable on the surface. On the other hand, system B, by convention, always represents fungal diseases by the joint use of FUNGI and the term DISEASES. If the indexing of system A is converted to that of B, via vocabulary reconciliation and without taking indexing conventions into account, inconsistencies and inaccuracies will inevitably occur.

A related problem is that of variations in the exhaustivity of the indexing in different organizations. Centre A may index with an average of 3 terms while B uses an average of 10. While it may be possible and useful to convert B's indexing to A's:

$$\left.\begin{array}{c}1\\2\\3\\4\\5\\6\\7\\8\\9\\10\end{array}\right\}\rightarrow\begin{array}{c}1\\4\\5\end{array}$$

since mapping from specific to general will usually be involved, the reverse procedure may be altogether more difficult.

Even spelling errors among index terms could affect the convertibility of data bases if these errors occurred frequently. In a study of 3600 terms drawn from 11 data bases, Bourne/48 discovered that, while index term misspellings were less than 0.5% in one source, they reached almost 23% in another. With inaccuracies as high as 23%, mapping from one vocabulary to another could be a rather haphazard process when terms are derived from the bibliographic records themselves. However, as Zamora/563 has pointed out, automatic recognition and correction of spelling errors may be possible.

The reconciliation of vocabularies seems possible, by means of an intermediate lexicon or in some other way. Whether such reconciliation will allow the useful merging of data bases, or the useful exploitation of a variety of data bases in a network situation, is open to considerable doubt. To achieve a more useful level of convertibility among data bases, taking indexing policies into account, is much more difficult (and costly) than vocabulary conversion. Moreover, it is not fully achievable since the policies/protocols of a particular organization may not be completely and explicitly stated. It is perhaps hardly surprising, then, that, while research on vocabulary convertibility has proceeded for at least 20 years, actual implementations of conversion or switching projects, in a real information service environment, are practically nonexistent.

It seems justifiable to conclude that the flurry of relatively uncoordinated activity in information systems development, evident over the last 20 years, has resulted in an undesirable proliferation of overlapping vocabularies*. Unfortunately, when a new information service is created, the tendency is to develop a new vocabulary rather than to adopt or adapt one that may already exist in the same general subject area. Consequently, when organizations become more serious about co-ordination, the sharing of resources, or the establishment of a network, a terminological barrier is immediately encountered: different vocabularies already exist and need to be reconciled.

This is not to suggest that in all situations there will be an existing vocabulary that will meet local requirements. Nevertheless, the fact that the proliferation of thesauri does not seem to be diminishing raises the question as to whether all newly established vocabularies are really necessary or whether needs could have been met by adopting or adapting some existing tools.

Although the desirability of vocabulary compatibility has been assumed throughout this discussion, not everyone accepts this assumption. Gaaze-Rapoport[/167], for example, takes a completely different view:

> When discussing problems associated with IRL's** one often speaks about so-called compatibility of IRL's and sometimes it is even required that the unified system of IRL's include only compatible languages. It seems to us that such stress on compatibility is unwarranted and in essence disorients both the developers and users of information service systems. Actually, by compatibility of languages is meant the possibility of reading a text written in one language by means of another language without a translator or translating program. This is possible only in the case of complete coincidence of the vocabulary and grammar of both languages, i.e., when actually there is one language, or when one of the languages is an expansion of the other, i.e., its vocabulary and grammar contain as a part the vocabulary and grammar of the narrower sublanguage. In this case the texts written in the narrower language (sublanguage) can be regarded also as texts of the broader supralanguage (but not vice versa). In all other cases there is no compatibility and it does not make sense to speak about it, and one must simply master both languages. It is senseless to speak, for example, about the compatibility of Russian and English languages or Chinese and Hindi languages. Analogously, it is senseless to speak about the compatibility of artificial languages, for example ALGOL and APL. Therefore, talks about the compatibility requirement create the illusion that such compatibility is possible in principle and, therefore, they are harmful...
>
> Thus, in our opinion, the problem consists not in attempts to create a system of compatible information retrieval systems, which are doomed to failure beforehand, but in the rational use

---

* This has occurred even within the UN itself. The Bibliography of United Nations Thesauri, Classifications, Nomenclatures[/41] lists, for example, four separate vocabularies in population and over 20 on various facets of agriculture, forestry and fisheries.

** IRL = information retrieval language.

> of the entire rich set of different IRL's and in selecting such a subset from this set which best meets the requirements of the specific subject field or specific problems facing a particular local information system. (Page 46)

Gaaze-Rapoport is correct, of course, in reminding us that compatibility should not be the overriding consideration in selecting an index language for a particular application. There seems little point in achieving compatibility with other organizations at the expense of using a tool less than ideal for the purpose intended. Nevertheless, the author does adopt a very narrow view of compatibility. According to this view, microthesauri would be considered compatible with their parent thesaurus, but the other approaches discussed in this section of the report would not be regarded as "achievement of compatibility". Clearly, the authors of this report have taken a wider view of compatibility by including also convertibility and by considering compatibility factors (e.g., structural compatibility) affecting convertibility.

4. SOFTWARE

Software includes both programs and data structures. Programs may be further subdivided into applications programs, which are written by users, and system control programs, which enable the computer to function and control its own operations (Saffady/[424]). The importance of system control programs can be illustrated by the following description of how a computer program is processed (Withington/[555]). Programs written in high level languages are processed by compilers into a form that the computer can execute. When the program is executing, an operating system allocates various hardware resources to the program as needed. When the program calls for data from files, access to the files is controlled by some kind of data base manager. When the user is located at a remote terminal, a communications control program is required to send and receive messages and device control instructions as well as special characters to acknowledge correct receipt of messages. Fosdick/[158] provides an introduction to operating systems and other systems support software.

Software compatibility deals with the use of a language that can be translated into the (perhaps very different) instruction sets of several machines (Meek[324]). A language acceptable to more than one type of computer is compatible. A program written in such a compatible language is portable if it produces acceptably similar results on two different machines. Portable software is compatible but the converse is not necessarily true. Compatible programs operating on different machines may produce the same results given the same data, but they may be machine-dependent in some way and therefore not portable. Software is upward compatible if identical data will produce the same results on a more recent version as on an older version even though the newer version may also accept additional forms of data.

4.1 Programs and programming languages

There is a hierarchy of programming languages that can be used on a given computer. Machine language refers to instructions written in machine code which can be immediately obeyed by a computer without translation. It is obviously machine-dependent. Assembly languages are at a slightly higher level, as they are symbolic languages which must be processed by an assembler in order to be converted to machine code. The assembler translates symbolic operation codes and addresses to machine language form and groups the resultant machine language program from its constituent parts. Assemblers of different computers are typically incompatible and programs in assembly language are not portable to other types of hardware. Assembly language is often used where the programmer requires close control over internal operations of the computer, such as for system control programs.

High level languages are even further removed from machine language, with each instruction corresponding to several machine code instructions. Compilers are the translation programs for high level languages. Programs written in high level languages are much more portable, for with only minor changes they can often be compiled to run on various machines. Williams and Schipma/[547] provide an illustration of this in their discussion of software development for a computer search centre. Their desire to develop programs for use at many installations led to the adoption of PL/I, a high level language. Between April and September 1969 their programs ran successfully at seven different computer installations on different models of the IBM 360 series, using different disk drives, and different versions of the operating system.

Even if a program is in a higher level language, problems may arise when computers have incompatible operating systems which impose different constraints on the programs they control. This is particularly true of microcomputers where the diversity of available operating systems is a barrier to the efficient use of these machines (Lu/[294]). Fortunately, alternative operating systems are easily implemented where appropriate circuit cards are available. For example, both TRS-80 and Apple microcomputers have compatible circuit cards for the CP/M operating system, which is widely used on other makes of microcomputers (Blair/[43]).

Sammet/[430] has described a number of factors that can contribute to incompatibility of programs and/or programming languages. A program will not be machine-independent if it makes reference to hardware unique to a particular machine. A program will not be compiler-independent if its compilation causes the program to produce different results depending on which compiler is used. Programming languages may have dialects, minor variations that will lead to incompatibility if features unique to a given dialect are used in writing a program. Subsetting and extension are also possibilities. If a program is written in a subset of a programming language, it should be upward compatible. Extensions introduce new features unique to a particular version of the programming language and are another source of incompatibility. Many of these sources of incompatibility are due to the difficulty in defining programming languages in a complete and accurate fashion. Thus, each compiler writer may make somewhat different decisions which subsequently lead to different dialects of a language.

Standardization has been used in an effort to achieve compatibility of several of the more widely used programming languages. But availability of a standard does not necessarily guarantee compatibility because parts of the standard may remain open to interpretation and many compilers are more restrictive. For example, Day/[107] notes that conformity with the Fortran standard is not enough. The subset that constitutes Compatible Fortran is sufficiently different from ANSI Standard Fortran to require a separate description. Day's Compatible Fortran is acceptable to most compilers so that programs written in it can be run on many different machines.

Wasserman/[526] has noted the decreasing reliance on highly machine-dependent programs, particularly those written in assembly language, due to the sharply lower costs of hardware and the availability of high level language processors for small as well as large computers. There are definite incentives for developing portable software (Brown/[56]): (1) economic grounds (much money is invested in software development so it is desirable for it to be reused); (2) hardware developments (with regular improvements in hardware, most sites are continually acquiring new hardware); (3) increased interest in networking. Brown's monograph is based on an advanced course on software portability supplemented with case studies of practical projects. Some of the techniques described therein are illustrated by the software that supports the Washington Library Network (WLN). As described by Brown and DeBuse/[58], certain conventions used in the design and implementation of WLN have made it portable to other sites: (1) top down design; (2) standard use of IBM environments; (3) structured coding techniques; (4) interfaces to the data base management system (ADABAS) and teleprocessing monitor (CICS); (5) standard naming and formatting; (6) use of a standard set of data structures and assembler subroutines to manipulate data; (7) identification of maintenance changes in programs. Furthermore the programming for online functions was as terminal-independent as possible. The WLN software has more than 250,000 lines of PL/I and IBM BAL code in more than 1100 program modules, running on IBM and IBM-compatible hardware. Sites where WLN software has been replicated include the National Library of Australia and the University of Illinois. Brown and DeBuse conclude that if a system is designed for ease of maintenance at perhaps some sacrifice of efficiency, it will be readily portable.

Walker and Harrison/[521] describe an automatic methodology for verifying program portability. In discussing portability, they distinguish between program conversion verification (testing that programs have been converted, data files moved, and documentation updated) and program quality verification (testing reliability, maintainability, and enhanceability of the converted system). The technology for verifying software portability has four parts: static analysis (inspect the structure of programs, files, data base and documentation), test case generation, dynamic analysis (verify the behaviour of the program against live, carefully selected test data), and quantitative program evaluation (compute statistics from data gathered during static analysis and dynamic analysis). Using their methodology, both program conversion and program quality can be verified.

The existence of computer networks means that, in principle, programs can be transmitted from one site to another. Turner/[491] has proposed an International ASCII Publishing Standard (IAPS) designed to eliminate the difficulties in transferring usable programs between different computer systems. It is a data transmission format which can be used for program or data base "publication" through the following process:

(1) a program or data base is developed using the software provided on one system;
(2) after development is complete, a special purpose conversion program is used to convert the format of the data or program into standard IAPS format;
(3) the file is transmitted to another computer;
(4) upon receipt of the transmission file, another customized conversion program is run to translate standard IAPS format to one acceptable to the second system;
(5) the data or programs are loaded and manipulated on the second system.

Videotex is a recently developed information transmission medium. One possible application of videotex is the transfer of software from a central computer to an intelligent videotex terminal. This has been termed "telesoftware" and raises portability problems if a program is incompatible with the processor to which it is transmitted (Storey/[469]). Sedman/[438] has described the features of a language called MicroCobol used in Prestel to enable a single version of a program to be transmitted to and executed on a range of different terminals. Based on COBOL, the language has been modified to make it suitable for interactive use on small machines and extended to include modern structured programming constructs. It has been found to be suitable for the intelligent business videotex terminal.

Brown/[55] cites system constraints that may lead to incompatibility, including differences in the peripherals available and in the memory size. A program may make certain demands on the computer environment that can only be satisfied by very few machines. For example, a program requiring a large number of peripheral devices or unusual and special equipment may not be portable. Similarly, a very large program may not be portable. In providing criteria for evaluating software for purchase, Garoogian/[171] points out the need for compatibility with both hardware (memory size, peripherals, and display requirements) and software (the available operating system and compilers). Armstrong/[14] presents descriptions of software for printed indexes, both commercial and noncommercial. Portability was one criterion for inclusion of packages in the survey. Each description includes details that could affect portability, i.e., type of computer, language used, memory requirements. The UNISIST inventory of software packages for information work will be even more complete, including software for a variety of information processing tasks. The National Center of Scientific and Technological Information (COSTI) in Tel Aviv is reponsible for collecting descriptive data on these packages, including factors that would affect portability.

A package intended to be portable needs to have far more written about it. Although there are no widely accepted standards for system documentation, some suggested standards have appeared. Enger/[143] presents a documentation standards manual developed for governmental and industrial clients. He outlines and discusses the contents of sixteen key documents that normally would be produced during design and implementation of a computer application. The documents are associated with six steps initiation (user request, user proposal, feasibility study, project plan), analysis (functional description, data requirements documentation), design (system/subsystem specification, data base specification, program specification), development (program documentation), implementation (test and implementation plan, test analysis report, user's manual), operation (operations manual, maintenance manual, evaluation report). Assuming that documentation standards written for batch operations are inappropriate for online applications, Rubin/[421] presents standards for online systems documentation. His comparison of elements in batch vs. online systems clearly illustrates the differences in system environment and thus the need for a different approach to documentation (page xiv):

Batch	Online
Preprocessing	Transaction logic
Keypunch instructions	Man-machine dialogue
Program flowcharts	Screen formats
File definitions	Input codes
Record layouts	Command syntax
Program documentation	Module documentation
Console operator documentation	Data base organization
Input forms	Data base access method
Output forms	Communications queues
Report distribution lists	Security tables
	Communications line control
	File directories
	Communication network layouts
	Terminal specification
	Terminal operator instructions

4.2 Command languages

One important software component of an online information retrieval system is the command language. Command languages can be viewed as special purpose, problem-oriented programming languages (Meadow/[322]). There is ample evidence that available command languages in various application areas have numerous differences. A number of comparative studies have been published with various objectives: (1) to provide command reference guides for multiple system users; (2) to facilitate comparative evaluation of features of different systems; or (3) to provide a basis for command language standardization efforts. Several have compared the command languages of bibliographic retrieval systems: Conger/[91] compares commands of the New York Times Information Bank, BRS, DIALOG, MEDLARS, DOE/RECON, ORBIT, and RLIN; Weiss/[528] compares BRS, DIALOG, ORBIT, and the Defense Technical Information Center DROLS; and Martin/[314] compares early versions of BASIS (Battelle), DATA CENTRAL (Mead), DIALOG, INTREX (Massachusetts Institute of Technology), LEADER (Lehigh), NASIS (NASA Lewis Research Center), ORBIT, RECON (Informatics), RIQS (Northwestern University), SPIRES (Stanford University), and STAIRS (IBM). Harper/[200] presents a table that gives a comparison of features of BRS, DIALOG, and ORBIT including their handling of truncation and word proximity searching. Auber/[16] suggests that differences among closely related languages (e.g., those with a common "ancestor") may cause more problems than differences among languages of totally different origin. He presents three comparisons: BRS/SEARCH (Data-Star) vs. IBM-STAIRS (Datacentralen); GRIPS/DIRS3 (DIMDI) vs. Eurolanguage (ESA/IRS); and DIALOG vs. QUEST (ESA/IRS). He found differences in print formats, command names, and available commands. In all these comparisons it is evident that, while there is a good deal of functional compatibility in the sense that each system enables the user to carry out similar functions, there are wide variations in details of the languages such as command names, syntax, and the use of special characters.

Although most comparative studies have dealt with bibliographic retrieval systems, other studies have looked at online cataloguing systems and online public access catalogues. Long/292 compared the message text formats used to initiate searches of systems operated by the Library of Congress (MUMS, SCORPIO), National Library of Medicine, OCLC, Research Libraries Information Network, Washington Library Network, Northwestern University, and University of Chicago. He looked at searches on personal name, corporate and conference name, title, author/title, subject heading, Library of Congress card number (LCCN), ISBN, and ISSN. Major differences were found to exist in formats for personal name, corporate and conference name, title, author/title, and subject. On the other hand, formats for LCCN, ISBN, and ISSN were quite similar. Differences in format were due to the different techniques used in searching these systems, which include truncated strings, keywords, and search keys.

Hildreth/219 has provided a very thorough comparison of ten online public access catalogues now operating in the United States: MELVYL (University of California), TLS--Total Library System (Claremont Colleges), BRS/SEARCH with a user-cordial interface (Dartmouth College), MSUS/PALS--Minnesota State University System: Project for Automated Library Systems (Mankato State University); ULISYS--Universal LIbrary SYStem (Mission College), NOTIS/LUIS--Northwestern Online Total Integrated System/Library User Information Service (Northwestern University), OCLC (Online Computer Library Center), LCS--Library Control System (Ohio State University), Maggie's Place (Pikes Peak Library District), RLG/RLIN (Research Libraries Group/Research Libraries Information Network). Classifying command capabilities in four categories (operational control, search formulation control, output control, and user assistance: information and instruction), Hildreth concluded that the comparison revealed more differences than similarities with a wide range of functions and search capabilities existing across systems. Definitions of Hildreth's categories illustrate the possible range of capabilities. Operational control includes functions performed prior to, during, or after the actual search process to facilitate search: logon, logoff, data base selection, output interrupt, error correction, request for news or explanations of data bases and commands. Search formulation control includes logical or matching operations on specified search terms to identify and isolate document descriptions pertinent to the inquiry. Output control governs display or printing of bibliographic references selected during the search process. User assistance provides guidance and instruction facilities. Although the command languages analyzed were designed to provide access to bibliographic records in library catalogues via traditional (and other) access points, little similarity exists across systems. Command languages can vary in command names and syntax, abbreviation rules, punctuation, special characters, and explicit/implicit entry and default values. Because no standards exist, the same command name may invoke entirely different functions in different systems or different command names may invoke the same function in different systems. An additional variation is due to the fact that in some systems commands may be entered by the user in ways other than typing command words (e.g., menu selection, question-answer, function keys).

Given this diversity, it is not surprising that the literature contains pleas for standardization of command languages. Atherton/[15] proposed that the command language used in interactive retrieval systems should become standard as far as terms and symbols used by the user were concerned. Otherwise as an individual uses different systems, he must perform mental conversions as he moves from system to system. The greatest progress along this line has been made in Europe, with the implementation of a common command language (CCL) by several hosts on Euronet-DIANE. The need for harmonization efforts to make it easier for the multihost user is well documented by Scharff/[432], who notes that mid-1981 Euronet-DIANE provided access to 250 data bases and banks, operated by 26 different hosts, loaded on ten different computer types, searched by thirteen different command languages. Furthermore, there are as many different login procedures as there are hosts. The implementation of CCL by several hosts is at least one step toward greater uniformity as perceived by the user. The development and implementation of CCL are described in greater detail in section 8.6. In the United States, ANSI Z39 Subcommittee G on Standard Terms, Abbreviations, and Symbols for Use in Interactive Information Retrieval has developed a working paper for review/[88]. Issues raised by the Subcommittee include: the need for compatibility between ANSI and ISO standards in this area, the desirability of including commands for both bibliographic retrieval systems and online catalogues, the possibility of standardizing menu-driven systems, and the need to standardize symbols for truncation, Boolean operators, and other items.

Negus/[361] observes that some people advocate the use of translating software as an alternative to standards such as CCL. Such translating software could be located either in a node serving as a gateway to communications networks or in an intelligent terminal. Either approach is limited, however, because a translation program must be done for each pair of languages (the one familiar to the user and the one for the host to be accessed). Any one implementation of this approach is likely to be limited to only a few such pairs. Furthermore, unless there is agreement regarding the features that all command languages should possess, it would not always be possible to completely translate from one command language into another. Negus argues persuasively that just as the interconnection of systems is not possible without standards at the signal level, standards should be applied at higher levels. If a standard command language were used in all hosts, one could still provide a tailored interface for a given user group. However, only one translating program would be required to convert this to the common command language for accessing the various hosts. This approach is consistent with the recommendations of the British Computer Society Query Language Group/[53]. They suggest that systems should be designed in such a way that different styles of user interface can be supported, via a basic command mode that would not necessarily be visible to the user. One system already supporting multiple command languages is 3RIP. As Edström and Wallin/[136] describe, 3RIP allows the user to introduce his private command language or to simulate the command language of another system.

It is not clear that a standard command language would necessarily be the best choice for all systems, however. In their feature comparison of an in-house retrieval system (MASQUERADE) with a commercial search service (ORBIT), Boyle and Miller/[50] caution that one must be careful not to attempt to duplicate a commercial system blindly in-house. The designer of an in-house system must take into account the applications to be performed, the operating environment, and user needs. Certain features of a commercial system, which tries to accommodate a range of queries to a wide variety of data bases, may simply be irrelevant to local needs. This contrast between in-house and commercial environments is similar to the contrasting needs of casual and experienced users. The casual user would like a language that is simple and easy to learn, while more experienced users would like additional features. Gebhardt and Stellmacher/[172] propose providing the user with a simple, clearly designed subset of a command language that nevertheless has all important functions. Additions to modify, shorten, improve, or extend it would be available to the experienced user. Meadow/[322] has suggested the possibility of machine assistance in tailoring the command language to individual users:

> ... it may be possible, using intelligent adaptive systems..., to design a system which can enable any user to begin productive use with a few minutes prior effort, and then to teach advanced methods of use gradually and painlessly, guided by a program that senses how well the user is performing at any time. The essence of this is that we must stop trying to design a single language which is somehow best for everyone. Instead, we must design interface programs that can help users, at whatever their proficiency level, deal adequately with their processing programs. (Page 321)

Both machine assistance and interface design are treated in detail below. Although standardization of command languages is clearly one approach to achieving compatibility in certain settings, translating software is also likely to play an important role, particularly when system designers try to accommodate the needs of a variety of users.

A type of command language of interest from the perspective of primary publication is that supporting text preparation and interchange. In a recent review of interactive editing systems, Meyrowitz and van Dam/[330] note that several ANSI and ISO committees are considering standardization of generic markup languages, text-processing languages, text formatting, text editing, and text structures. Although they feel that it is too early to agree upon a standard editor, they advocate acceptance of a standard reference model for editing which can serve as a base for further research and development. Le Roux/[283] has outlined ISO/TC 97/SC 18 concerns in this area, including: a) assessment of user requirements for preparation, storage, retrieval, editing, formatting, message processing, and the transmission of text; b) interworking between different systems and services (e.g., facsimile, text transmission via conventional online networks, videotex, teletex); c) command languages for text preparation and interchange. Although the attention of information system designers has to date been focused primarily on command languages for retrieval, it is evident that future information systems must be concerned with the compatibility of languages for text generation as well.

4.3 Data base management systems

Data base management systems (DBMS) are sets of programs that allow a user to create files, to maintain them by adding, deleting and changing records, and to retrieve stored information. The DBMS may have a data definition language for defining data bases and a query language for retrieval. In comparing two such complex software packages, there are a number of different levels of compatibility (Valantin/[504]):

(1) data interchange--this is the lowest level and can be considered both in terms of the compatibility of physical structures and of logical structures. The physical structure is the data container. For physical structure compatibility, the DBMS should follow standards such as the international standard for bibliographic information interchange, ISO 2709. Logical structure refers to the data elements included in the bibliographic record and their form. Full compatibility may not be required if formats allow easy transformation of data elements.

(2) functional compatibility--two systems are functionally compatible when users can store and manipulate essentially the same kinds of data elements and can perform in one way or another the same classes of operations on the data.

(3) end user compatibility--with full compatibility at this level, a user need not know with which system he or she is interacting as all commands and command sequences are identical. This can be achieved through use of a high level network command language. Full compatibility between systems would require a combination of all three forms of compatibility.

Valantin's thorough study comparing the functional compatibility of CDS/ISIS and MINISIS can serve as a model for future studies of this type. CDS/ISIS is the Computerized Documentation System which runs on IBM 370, RIAD and AMDAHL mainframe computers and is maintained and distributed by Unesco (Pobukovsky/392). MINISIS was developed by the International Development Research Centre in Canada for operation on the Hewlett-Packard HP3000 range of minicomputers and it likewise is available for distribution (Boyce/49). According to Valantin, Unesco has already made CDS/ISIS available to more than 50 national and international organizations and IDRC has provided MINISIS to 18 institutions including nine in developing countries. Both systems are based on the ISIS (Integrated Set of Information Systems) software originally developed by the International Labour Organization as a generalized information storage and retrieval system with particular emphasis on bibliographic applications. Systems in the ISIS family are all characterized by certain common features: similar data base structure, similar record structure (leader, directory, variable length data), similar data structure (variable length fields and subfields), generalized data base definition, online data entry/modification and retrieval, and a variety of other basic functions such as print formatting. All use the ISO standard 2709 format for data interchange, making it possible to exchange files among installations.

The evaluation of functional compatibility was undertaken because both systems attempt to provide generalized data base management services and are members of the ISIS family. Information on any differences would enable them to achieve functional compatibility as new releases of the software are written. In addition the study provides a useful checklist of features that could be applied to other software packages to determine their suitability for different applications. A function can be thought of as an operation that a user can perform or have performed on the data base or the data contained therein. The evaluation identifies the available functions and where possible the specific commands or set of commands associated with them. The findings are presented as tables comparing such functions as data base definition and creation, online data entry, online data modification, data integrity checks/data validation, online record status changes, online retrieval, inverted file creation and maintenance, element extraction and sorting, printing and print formatting, ISO format conversions and data base restructuring, file maintenance, and security. Valantin concluded that there was a very high degree of functional compatibility between CDS/ISIS and MINISIS. The main differences were attributable to differences in the system control programs provided by IBM and Hewlett-Packard and to different philosophical and theoretical approaches to system design. In particular, MINISIS is based on a clear conceptual framework, the relational data base approach. Functional compatibility does not always correspond to user compatibility, however. For functions with direct equivalents in the two systems, there is a moderate degree of compatibility in command names at the end user level. Valantin recommends that synonyms be provided for the most commonly used commands so that users can feel familiar with both systems.

A recently completed study by Park and Stearns/383 evaluates the IMD IV+V system software package (developed at the Institut für Maschinelle Dokumentation in Graz, Austria) for UNISIST *Reference Manual* implementations. Although they did not undertake a detailed functional compatibility study comparing IMD IV+V to

CDS/ISIS and MINISIS, they do suggest that it is likely to correspond more closely to MINISIS than to CDS/ISIS because it is based on the relational data base model. Park and Stearns also comment on the portability of IMD IV+V, observing that because the package is written in PASCAL it is likely to be portable to a variety of machines. Unlike other widely used programming languages that have machine-specific extensions, PASCAL is identical across many different computers thus supporting program portability.

The likelihood of achieving compatibility in the area of archives automation is increased through the use of software packages for archives and records management applications. In the United States two projects, one based on mainframe computers and the other on microcomputers, illustrate the capabilities that such software packages can have. SPINDEX (Selective Permutation INDEXing) is a group of programs developed by the National Archives and Records Service (Sahli[426]). It is machine dependent, relying on the availability of IBM 360/370 computer hardware, but it is otherwise fully transportable and it has been implemented by a number of organizations. SPINDEX components include programs to support input and file management, generation of archival registers, and index production. Limitations of SPINDEX include its machine dependence and the reliance on expensive computer hardware unavailable to many organizations. To offer an alternative, MARS (Microcomputer Archives and Records Management Systems) has been developed (Kesner and Hurst[262]). Written in BASIC, it is portable and inexpensive. It is also interactive, a capability not supported by most SPINDEX installations. The use of such software packages to create records at different installations establishes a basis for sharing records of archival holdings.

The relational data base model characteristic of MINISIS and IMD IV+V is one of three commonly selected as the basis for the design of a DBMS. The other two are the hierarchical and network models. Salton and McGill[429] provide a description of all three together with examples. Michaels et al[331] compare and contrast the relational and network approaches in terms of data definition, data manipulation, data protection, data independence, and system performance. They conclude that since such systems are to serve the needs of diverse communities of users, no single approach to data base management seems either desirable or likely to emerge as dominant. Thus it is unlikely that one model would be selected as a standard to the exclusion of others. Although there may be functional compatibility between data base management systems based on models of a different type, such systems may not be end user compatible.

Even compatibility in file structure does not guarantee identity in retrieval performance. Systems for bibliographic retrieval commonly employ inverted files to maintain indexes to record locations. But the content of these inverted files is not necessarily the same because decisions must be made as to which fields are to be stored in the inverted files and how they will be stored. A study by Rouse and Lannom[419] reveals a number of differences in the inverted files maintained by BRS, DIALOG, and ORBIT for the same data bases (e.g., ERIC). Different fields were selected for inclusion in the main inverted file (e.g., ORBIT did not include terms from the abstract at the time of the study) and field values were entered in different ways (e.g., a multiword thesaurus term such as "information retrieval" could be entered simply as a phrase in the inverted file in one system and as both a phrase and individual keywords--"information" and "retrieval"--in another). These differences in file "loading" frequently lead to different retrieval results.

Levine[285] has observed that differences in storage organization may result from differences in system priorities. Inverted files allow fast response times but are time consuming to update; many systems emphasize ease of maintenance rather than fast response times for a variety of queries. Furthermore, many, unlike CDS/ISIS and MINISIS, cannot easily handle files of bibliographic data. They may lack the capability to handle variable length records, variable length and variably occurring fields, subfields, and long descriptive abstracts. Yet, they may be perfectly adequate for standard DBMS applications such as inventory and personnel files. Thus differences in demands of the application can lead to incompatibilities in the supporting software.

New developments in data base management systems are likely to result from new application areas, particularly networking. Nahouraii et al/^{347} describe an approach to data communication between different generalized data base management systems, valid for different logical structures in different types of systems on different computers. The user of one DBMS may thus retrieve records or field values from another data base even though the two management systems involved are incompatible. The communication technique relies on a store of information about how data are structured in a given DBMS. A similar concept has been introduced by Cousins and Dominick/^{94} who describe the "meta-base", a data base which has as its entities the attributes of other data bases. The meta-base contains complete definitions of all data items, all procedures that operate on these data items, and all access paths used to locate values of data items. One would search the definition of a data base to ascertain what types of data are contained within it prior to searching the data base itself to retrieve specific values of data items.

Lowenthal/^{293} suggests that data base systems designed for local area networks must surpass the capabilities of traditional data processing-oriented systems. In an office environment a DBMS should be able to handle both formatted records and "streams" of text, digitized images, or graphics. Both types of information should be able to coexist in the same data base and new ways of combining them should be possible. Lesk/^{284} gives examples of a number of such combinations. Starting with files of the AP news wire with current national and international news, the National Weather Service wire with United States weather observations, a map of the United States, a list of United States cities and towns with locations, and a list of United States airports, it is possible to have such services as: (a) draw the map of any area; (b) report the weather anywhere (by determining the nearest weather station and reporting its forecast); (c) find the closest airport to a town; and (d) find the closest news story to a town.

Although data base management systems are sufficiently complex software packages that it is unlikely that an entire system would be adopted as a standard, there are still possible areas of standardization within the DBMS field, e.g., programming language additions for data manipulation, data definition language, query language, device and media control language, operating system interface language, network interface language and protocols for message switching (Sibley /^{445}). There has been little progress to date in most of these areas. Without standards the conversion of files from one DBMS to another may be a laborious process because no generalized program for conversion can be written (Berg/^{36}). Yet such conversions are frequently undertaken as new systems with a greater range of features become available or as new hardware is acquired by an organization.

4.4 Machine assistance for searching

The availability of comparatively inexpensive microcomputers and intelligent terminals has led a number of researchers to investigate techniques for machine assistance to the users of online systems. One important function of such machine assistance is to shield users from the incompatibilities that exist among various systems and data bases. This approach has been referred to in a variety of ways, e.g., a user interface box, a front-end, a translating computer interface, a program for machine-mediated searching, a computer intermediary, an automated expert assistant. In each case the intent is to overcome barriers to the effective use of available retrieval systems (Doszkocs et al/[126]). In particular such an approach should allow end users to do their own searching (Dayton et al/[108]).

Williams/[548] has provided a useful classification of search enhancements, creating three categories: essential features, useful extra facilities, and more ambitious features. Essential features include: (1) use of permanently stored messages for automatic logon to a system using only a few key strokes; (2) preparation, checking and editing of statements in local memory before contacting the online system, saving online connect time and minimizing online errors; (3) storage of text locally (e.g., terms used frequently); (4) storage of output from the system. Useful extra facilities include: (1) manipulation of output; (2) floppy disks for more local storage. More ambitious features include: (1) input checking (check validity of commands and provide additional guidance for the user); (2) automatic selection of the optimum data base for a search from specification of the topic; (3) interfaces to allow queries in natural language; (4) automatic query modification. In a later paper Williams/[549] suggested some additional facilities: (1) transform data obtained from one system into the form required for input to another; (2) provision of help dialogues; (3) translation from one command language to that used by another system; (4) translation of search commands or output to the natural language of the requester; (5) dialogue to allow unskilled users to formulate searches by answering questions posed by the computer, with the search then carried out automatically on their behalf.

Williams/[548] suggests that the grouping of enhancements into essential features, extra facilities, and more ambitious features corresponds to a hierarchy of equipment which would be needed to implement them. The essential features could be made available on a microcomputer in the $2000-$4000 range. The extra facilities would require a larger scale microcomputer system, while the more ambitious features would have to be supported by a minicomputer or mainframe. In addition to computing power requirements, consideration of user needs will influence whether a given feature is implemented in a personal microcomputer, a local minicomputer, or the mainframe computer where the data bases reside. For example, it may not be feasible to maintain a data base directory in the microcomputer. Such a guide could be stored in a minicomputer on the network and shared by a number of users. However, the user might store information about certain frequently used data bases in his own microcomputer.

Although most work in the area of machine assistance is still experimental, certain devices are now commercially available. Both experimental and commercially available enhancements are surveyed here briefly to provide an indication of the range of techniques which has already been implemented. A detailed description of CSIN (which includes an intelligent terminal product as a spinoff) appears in section 3.8, so no further mention will be made of it in this section.

In recognition of the difficulty of logging on to hosts on Euronet, a Euronet-DIANE user interface box has been designed/[146]. Normally logon could require typing as many as 30 digits consecutively, with a typing error leading to an incomplete call. The box contains a programmable microprocessor which allows the user to enter and store access protocols. A battery is used to maintain protocols in storage for several hundred hours after the device is unplugged. It is a low-cost device which connects between the terminal and the modem. Williams' USERKIT (User Search Enhancement and Retrieval Kit for Interfacing to Terminals) is slightly more sophisticated/[550]. USERKIT is a microcomputer with programs permanently written into ROM and housed in a box which is interposed between the terminal and modem or acoustic coupler. It supports stored logon messages, offline preparation and editing of searches, and one key transmission of stored messages.

The possibility of using translating software to convert from one command language to another was cited earlier as one means of dealing with incompatible command languages. A number of researchers have developed such software to convert from a language which they hope is easy for a user to employ (such as natural language) to the command language of the system to be searched. Litvintseva and Senin/[289] describe the dialogue language processor in DILOS (Dialogue Information LOgic System) which converts natural language input into expressions that can be processed by the system. Fayen and Cochran/[150] have developed a microcomputer-based interface to an online catalogue data base resident on the BRS system. As the BRS command language cannot be used without training, the interface allows users to enter search requests in natural language form which the interface then formats properly for submission to BRS. Doszkocs and Rapp/[125] describe CITE (Current Information Transfer into English), a natural language interface to MEDLINE. As the interface program is implemented at the National Library of Medicine rather than in a remote intelligent terminal, the process is not simply one of translation. The software searches for documents that contain all or most of the key terms by using a special algorithm based on a combination of term weights. Retrieved records are ranked according to relevancy and the searcher can select documents of interest (relevance feedback) and command the system to expand the search by finding other items that contain terminology similar to selected citations. Because the initial query is expressed in natural language (English), the CITE system offers the potential for automatic query transportability to other data bases in MEDLARS.

While CITE uses relevance feedback to process queries on the mainframe computer, two other experiments illustrate the use of an intelligent terminal for this purpose. Morrissey/[340] describes a system in which the query is entered as a list of keywords. The interface program weights search terms, constructs commands in the Common Command Language for submission to DIMDI over Euronet, ranks output, and uses relevance feedback to refine the query. Jamieson and Oddy/[252] characterize their system interface as a "parasitical microsystem" which, given relevance judgments, updates the query by adjusting term weights and adding or removing query terms. These investigators cite the advantages of the intelligent terminal environment which allows one to experiment with alternative search mechanisms at reasonable costs.

Projects that expect the user to employ a common command language that is then translated into the command language of various systems include CONIT and MESSIDOR. CONIT (Conversion for Network Information Transfer) is a translating computer interface which both translates from the common language to the commands of the appropriate system (DIALOG, ORBIT, or MEDLARS) and transforms system responses into a common form before sending them on to the user. The network of different systems is thus made to look like a single "virtual" system to the user. The interface also provides some instruction. Marcus and Reintjes/[309, 310] describe the design and evaluation of CONIT while another paper by Marcus/[308] characterizes search aids that have been embedded in the interface such as a form of automatic data base selection, an advanced search save and multiple data base/system search regeneration and repeat facility, and automatic search on natural language phrases using keywording, stemming and Boolean combination. In contrast to CONIT, which links the user to one system at a time, MESSIDOR allows simultaneous search of several bibliographic data bases which may be on different systems (e.g., ESA/IRS, Télésystèmes) with different command languages. According to Moulinoux et al[342], MESSIDOR is a microcomputer-based system that treats multiple data bases on multiple systems as a distributed data base. Each query is automatically translated from a common command language and processed against each selected data base simultaneously. Results are received and stored, then reformatted to become the result of a distributed data base search. At the conclusion of a search MESSIDOR automatically logs off all systems and networks that have been used. For example, a search could be processed on NTIS, INSPEC, and PASCAL at ESA/IRS reached through Euronet and on EDF-DOC and TELEDOC at Télésystèmes reached through TRANSPAC.

Borrowing a concept from work in artificial intelligence, several investigators have characterized their work as the development of an "expert system" to assist the user in the search process. According to Pollitt[393], an expert system is a computer system that embodies organized knowledge concerning some specific area of human expertise sufficient to perform as a skillful and cost effective consultant. He suggests that, in the domain of online searching, an expert system may have knowledge in four categories:

(1) system knowledge--command language and facilities available in a search system from logging on and submission of search statements to printing of references and abstracts;

(2) searching knowledge--relating to strategy and tactics to be employed in a search;

(3) subject knowledge--one type would be term relationships and scope rotes as found in a thesaurus;

(4) user knowledge--characteristics of individual users such as previous searches and preferred journals.

Considering specifically the possible role of an expert system in searches for information on cancer therapy, Pollitt suggests that the system could also establish links to other types of files such as cancer registries or patient records as they become available to search online. Although Pollitt's work is in its initial stages, Meadow et al[323] have already implemented and evaluated IIDA (Individualized Instruction for Data Access). Through computer-assisted instruction IIDA can instruct users in how to search the DIALOG system and then assist with performance of the search by providing a diagnostic analysis of users' performance as well as answering questions about how to use system commands. Toliver[484] describes in more detail the expertise embedded in IIDA which allows it to detect both syntactic and strategic errors. The IIDA software has been further developed and implemented on a microcomputer for distribution as a commercial product under the name of OL'SAM (OnLine Database Search Assistance Machine). Franklin Institute Research Laboratory has done the development work which enables OL'SAM to access data bases on DIALOG, ORBIT, and MEDLARS through a single command language (Toliver[483]). It also provides computer-assisted instruction, assists in selection of appropriate vendors and data bases, manages the logon protocol, identifies syntactic and strategic errors, and helps the user recover from errors. As with the USERKIT, it allows composition of a search prior to the online session.

Two other expert systems have been called EXPERT and FRED. Marcus[307] describes EXPERT as an alternative approach to CONIT. CONIT teaches the user the command language and suggests search techniques. The user then needs to decide for himself how to apply these techniques using appropriate commands. EXPERT, on the other hand, takes charge of much of the search strategy formulation and reformulation through a series of questions directed to the user who responds on the basis of a menu or fill-in-the-form mode. FRED (FRont End for Databases) has three subsystems: (1) user interface subsystem that processes user entries, generates and transmits responses and prompts to the user, and delivers the requested information; (2) data base interface subsystem which sets up communication with the target data base, formats and transmits commands and queries to data bases, and handles returned information and messages; (3) knowledge base that contains facts and heuristics about users and available data bases (Crystal and Jakobson/99). Both of the interface subsystems share and act on information in the central knowledge base.

This brief survey of hardware/software systems providing machine assistance illustrates a continuum of "machine intelligence" with respect to the search process from the user interface box storing protocols to the expert system with knowledge about systems, searching, subjects, and users. All of these approaches represent attempts to assist the user confronted with incompatibilities when doing multiple system searching. Instead of requiring that the user remember the various details of different data base formats and command languages, they are stored in the interface and the differences are "transparent" to the user (Williams/544). Yet, as the discussion of command languages pointed out, even the expert system approach will bog down if the differences are myriad. Thus efforts to increase compatibility of systems through standardization are likely to benefit not only the experienced human user, but also the builder of expert systems. And with the proliferation of expert systems and other forms of machine assistance, there is the possibility of more searching performed directly by end users.

4.5 User-system interfaces

The various forms of machine assistance described in the previous section are all attempts to refine the user-system interface. This interface must be a focus of interest in a study of factors affecting information system compatibility because, as Hildreth/219 has stated:

> The barriers to effective communication at the man-machine interface are becoming widely known, and cognitive incompatibility between the user and the system is a real and present deterrent to user acceptance of online catalogs. (Page 41)

This "cognitive incompatibility" is a more nebulous concept than the other types of incompatibility discussed in this report, but it is an area that has been receiving increasing attention in system design for the last few years. One reason is economic. As Branscomb/51 observes:

> End-user satisfaction is rapidly becoming a decisive competitive factor. That is why the information industry is paying increasing attention to a field of applied psychology called human factors science and engineering, or ergonomics. (Page 757)

Hildreth has suggested that the overall structure of the user-system interface may be divided into the following parts or components:

(1) physical component--e.g., input/output equipment, physical structure of the workstation;
(2) organizational component--e.g., institutional setting;
(3) personal component--the user;
(4) communications component--dialogue modes and techniques, language of interaction (commands, prompts, messages, etc.);

(5) functional component--features providing operational control, search formulation control (including access points), output control, user assistance.

Figure 6, taken from Hildreth, illustrates some of these components and in particular enumerates some of the variables that can be used to characterize system users. The major activities and processes which take place at the interface are listed in the shaded area.

With the increasing attention to end user satisfaction, system designers have tried to identify attributes that make a system interface "user-oriented," "user-friendly," or "user-cordial." Dehning et al/110 have reviewed definitions of "user-friendliness" in the computer science literature and have found no consensus about its meaning. Those definitions that do exist are often not sufficiently detailed to provide any concrete guidelines for system design. They comment that:

> ... this is no surprise, because user-friendliness means user-orientation, and system design has been system-oriented until now. (Page 24)

At least part of the reason for a lack of consensus is that "'user cordiality,' like beauty, is in the eyes of the beholder" (Matthews/315, page 1070). It is much easier to design a system that will be operated by a homogeneous group of people with equivalent knowledge, training and experience, than a system for a heterogeneous population. The users of many systems are diverse and an interface which one group might term "user-friendly" could be regarded by another group as "user-hostile". Even "naive users" can be further subdivided, as naiveté can take many forms (Fox and Palay/162):

computer naiveté: the person has no experience with a computer;

interface naiveté: many information system interfaces are complex and idiosyncratic, requiring many hours of training;

information structure naiveté: how information is structured in the data base may be different from the user's view of the structure. Hence the ability to specify an information request using the appropriate commands does not guarantee that available information will be found.

Any of these types of naiveté can hinder successful use of a system, so information systems should be designed to remove these impediments.

The need to implement a single interface has been removed by new developments in hardware and software, as suggested by the earlier discussion of machine assistance. Rather than trying to optimize a single interface for a large heterogeneous user population, more than one interface can be implemented. Goldstein and Ford/182 have presented three alternative configurations for multiple interfaces: in intelligent terminals at the user site, as front-end processors at a central site, or as software modules in the main computer system (see Figure 7). They distinguish between systems operations and the interface as follows:

> All bibliographic retrieval systems support, nominally, the same set of operations. The user interface is, in reality, a matter of format; this includes the choice of command language as well as display formats. Hence the different interfaces are the result of subjective decisions by system designers as to the best format for the end user.

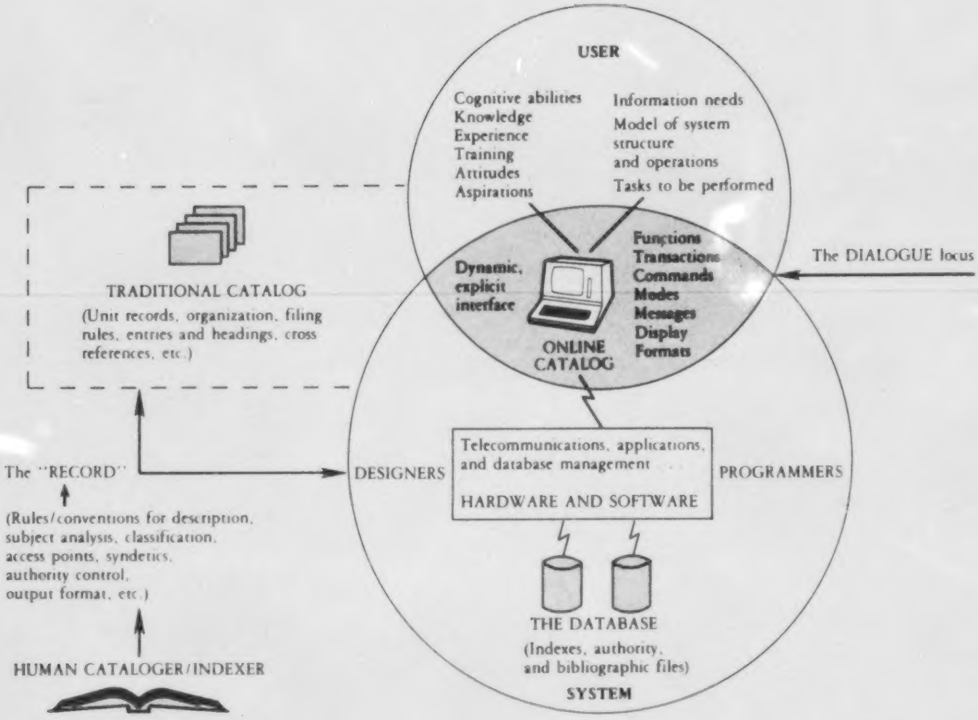
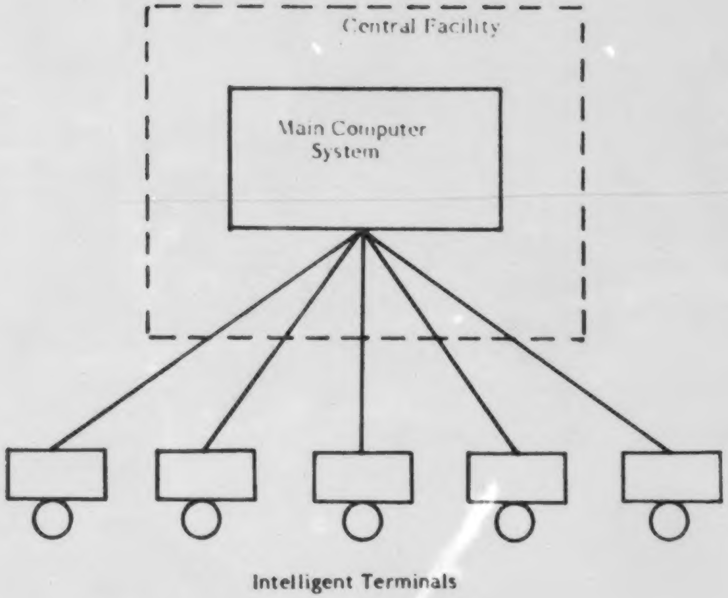
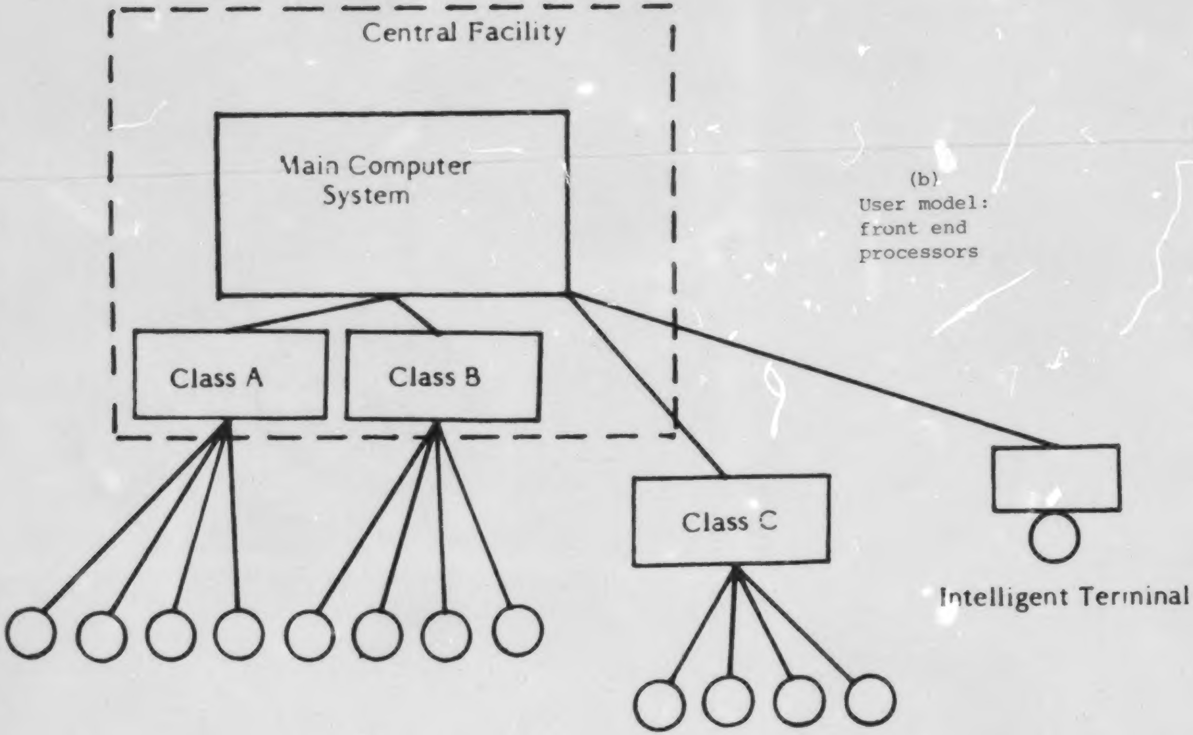

USER
Cognitive abilities
Knowledge
Experience
Training
Attitudes
Aspirations
Information needs
Model of system structure and operations
Tasks to be performed
TRADITIONAL CATALOG
(Unit records, organization, filing rules, entries and headings, cross references, etc.)
Dynamic, explicit interface
ONLINE CATALOG
Functions
Transactions
Commands
Modes
Messages
Display
Formats
The DIALOGUE locus
The "RECORD"
(Rules/conventions for description, subject analysis, classification, access points, syndetics, authority control, output format, etc.)
HUMAN CATALOGER/INDEXER
DESIGNERS
Telecommunications, applications, and database management
HARDWARE AND SOFTWARE
PROGRAMMERS
THE DATABASE
(Indexes, authority, and bibliographic files)
SYSTEM

Figure 6. The user-system interface in online information retrieval Reproduced from Hildreth/219 by permission of OCLC.


Central Facility
Main Computer System
Intelligent Terminals

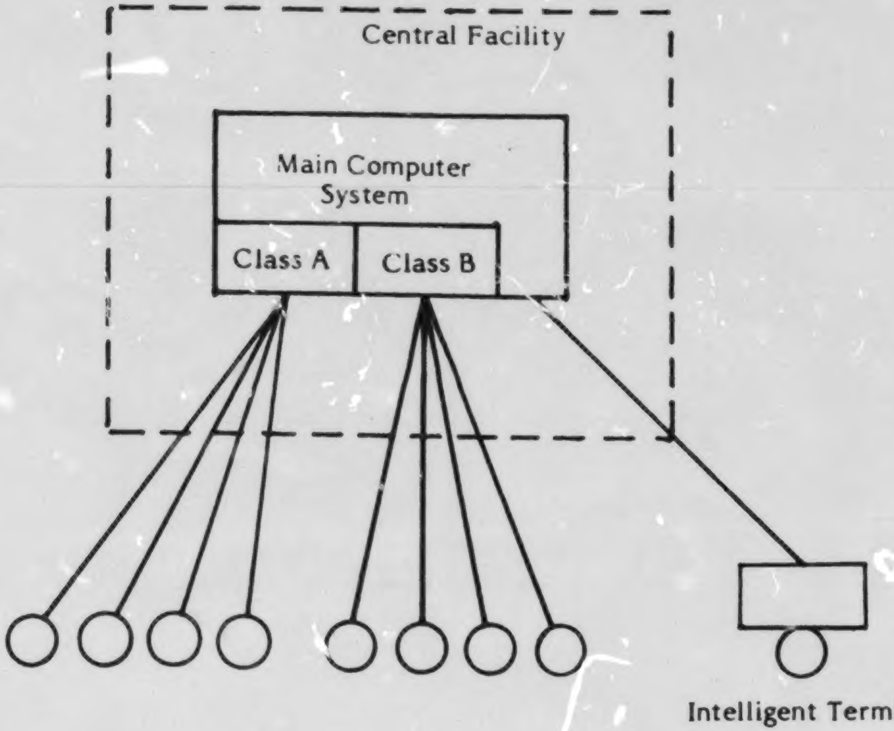
(a)
User model:
intelligent
terminals

Figure 7. Alternative configurations for user-system interfaces. Reproduced from Goldstein and Ford/[182] by permission of the publisher of Online Review, Learned Information Ltd.


Central Facility
Main Computer System
Class A
Class B
Class C
Intelligent Terminal

(b)
User model:
front end
processors

Figure 7 (continued). Alternative configurations for user-system interfaces. Reproduced from Goldstein and Ford/182 by permission of the publisher of Online Review, Learned Information Ltd.


Central Facility
Main Computer System
Class A
Class B
Intelligent Terminal

(c)
User model:
software modules

Figure 7 (continued). Alternative configurations for user-system interfaces. Reproduced from Goldstein and Ford/182 by permission of the publisher of Online Review, Learned Information Ltd.

> The technology, today, allows us to separate the retrieval capabilities (operations) of the central system from the format (interface) presented to the end user. (Page 275)

This distinction between operations and interface parallels Valantin's distinction between functional compatibility and end user compatibility. There is an apparent contradiction, however. When talking about end user compatibility, Valantin stated that it is of concern if a particular user will access several systems, because end user compatibility eliminates the need to learn several sets of commands and access protocols. Yet in the context of a single system "end user compatibility" for a heterogeneous group of users may imply the need for multiple interfaces, each "cognitively compatible" with some subgroup. These are evidently conflicting goals unless each system offers multiple interfaces which are compatible (e.g., a standard "experienced user" interface and a standard "naive user" interface).

The absence of available standards (or even guidelines) for user-system interfaces has already been noted. Nevertheless, some progress has been made in developing guidelines for portions of the interface. In the case of displays, for example, there is a growing body of research in cognition that can be helpful in understanding how displays should be formatted (Burns/65). Hickey and Spies /218 cite a number of guidelines and standards for the video display of information. Singleton/447 presents a display design checklist and Engel and Granda/142 have compiled guidelines for display formats, frame content, command language, error prevention and recovery, and response time. The guidelines are based on the application of known human factors and psychological principles, analysis of reported experimental studies in various behavioural disciplines, and examination of IBM customer and user comments. Stibic/466 presents a list of general properties of user-friendly online systems, noting that the requirements of experienced users on the one hand and of beginners or casual users on the other hand are frequently contradictory. The beginning user may prefer an interface that is simple, explicit and of maximum help while a more experienced user frequently prefers short command codes, brief messages, and flexibility. In spite of this disparity, Stibic offers some suggestions for making systems more friendly even for heterogeneous user populations. Mehlmann/325 presents an approach to developing an interface for systems to be used by "non-data processing" people who must work with terminals. The objective is to develop an interface that will make the system a tool helpful in the user's work.

Existing guidelines must be used with some caution. As Maguire/298 points out, there is a tendency for recommendations based upon a limited field of experience to be offered as more general advice. This can lead to conflicting guidelines from different sources. Recognizing that the applicability of certain guidelines may be situation-dependent, the system designer should consider each suggestion as an alternative strategy to be applied only in particular situations.

There are also situations where no guidelines exist either because the necessary studies have not been carried out or because new types of systems are being designed. Shneiderman/444 has written a book in which he advocates controlled, psychologically-oriented experimentation on problems of computer and information systems design. He describes appropriate research methods and includes a number of "researcher's agendas"--system design questions in need of further research. As one example of such a research study in a library context, Geller and Lesk/174 describe a comparative evaluation of menu and keyword retrieval systems as the means of access for a library catalogue. In the area of new types of information systems, one can turn to reports of case studies as one source of guidance. Simon and Doszkocs/[446] describe the design of an interface for a numeric/textual data base, the Laboratory Animal Data Bank. Because it was to be accessed by scientists without formal training in retrieval as well as by trained intermediaries, a decision was made to provide both a direct (command-oriented) search and a structured (menu-oriented) search. In addition to handling search strategy formulation, the interface supports report generation and statistical manipulation of retrieved data. In library environments it is the arrival of online public access catalogues that has focused attention on user-system interface design. Berger and Klemperer/[37] briefly describe the online catalogue developed by the Division of Library Automation of the University of California to serve nine University campuses. They note a number of aspects of interface design intended to make the system "friendly": separate dialogue modes (i.e., lookup vs. command) for different types of users, an extensive help/error system, and modifications of the terminal keyboards to make them less confusing.

With the increasing interest in serving the needs of heterogeneous user groups, it seems unlikely that, at least in certain application areas such as online public access catalogues, there will be a single standard interface. But there is the possibility that as studies provide a greater understanding of factors that can contribute to a system's "user-friendliness", standard levels of interfaces could evolve. In this way a given system could accommodate different types of users and the user would find a compatible interface when he moved to another system using the same levels.

4.6 Problems relating to software

Because hardware costs are decreasing while the costs of developing software for new applications continue to be high, portability is an important property of computer programs. Higher level languages which can be translated into the perhaps very different instruction sets of several machines enhance portability. Standardization of higher level languages has been accomplished for a number of the more widely used programming languages. Nevertheless, programming languages may have dialects, minor variations that will lead to incompatibility if features unique to a given dialect are used in writing a program. Differences in computing environments, including both available hardware (e.g., memory size and display devices) and software (e.g., operating system and compilers), can also limit program portability. Compatible programs operating on different machines should produce the same results given the same data, but they are not portable if they are machine-dependent in some way.

In the development of information systems one is concerned with command languages, data base management systems, and the user-system interface. As Vaskevitch/[510] has suggested, standardization in any one of these areas is a good indication that the facilities required by users are well understood. This is illustrated by the development of a common command language (CCL) to access multiple hosts on Euronet. Although it has been possible to identify and standardize a set of commands for bibliographic retrieval, other application areas still display considerable diversity. Command languages can vary in command names and syntax, abbreviation rules, punctuation, special characters, and default values. An additional variation is due to the fact that in some systems commands may be entered by the user in ways other than typing command words (e.g., menu selection, question and answer, function keys). Neither languages for text preparation nor for access to online catalogues are yet standardized. In addition, as will be described in chapter 6, there has been little progress in developing a standard command language for the retrieval, manipulation, and display of numeric data. Standards are likely to be developed only if a consensus can be reached regarding user needs in each of these areas.

Where standards are unavailable, there is still the possibility that functional compatibility can exist; each language or system supports the same class of operations. If two languages or systems are functionally compatible, it should be possible to develop translating software as an alternative to standards. However, to the extent that there are features unique to each language or system, a translating program can only be partially successful. Data base management sytems (DBMS) are sets of programs that allow a user to create files, to maintain them by adding, deleting and changing records, and to retrieve stored information. Two data base management systems are functionally compatible when users can store and manipulate essentially the same kinds of data elements and can perform in one way or another the same classes of operations on the data. Although available data base management systems may have a high degree of functional compatibility, they generally lack end user compatibility in that they require mastery of different commands and command sequences. Due to the lack of standards the conversion of files from one data base management system to another is often a laborious process because no generalized program for conversion can be written.

The availability of comparatively inexpensive microcomputers and intelligent terminals has led a number of researchers to investigate techniques for machine assistance to the users of online systems. One important function of such machine assistance is to shield users from the incompatibilities that exist among various systems and data bases by having programs in the microcomputer or intelligent terminal keep track of these differences. This approach can work effectively, however, only if there are not large numbers of differences to be accommodated. Design of the user-system interface is receiving a growing amount of attention because "cognitive incompatibility" between the user and the system is a deterrent to user acceptance. With the increased attention to end user satisfaction, system designers have tried to identify attributes that make a system interface "user-friendly". There is a dearth of standards (or even guidelines) for user-system interfaces. There is also a tendency for recommendations based upon a limited field of experience to be offered as more general advice, which can lead to conflicting guidelines from different sources. With the increasing interest in serving the needs of heterogeneous user groups, it seems unlikely that there will (or should) be a single standard interface.

5. HARDWARE

Computer hardware includes several classes of components (Hellerman and Smith/208):

(1) transformation devices are the circuits that actually carry out processing;

(2) transducers (input/output devices) change information from one physical form to another and hence serve as communication links between the computer and its environment (e.g., card readers, terminals, printers, plotters);

(3) storage devices handle both data and instructions and vary in capacity, access speed, and cost (e.g., tape, disk, drum);

(4) communication and control devices are circuits, including networks or busses, that direct the flow of information between other functional parts of the hardware system (e.g., input/output channels control the flow of information between transducers, auxiliary storage devices, and main storage).

This classification is followed below with discussions of compatibility as it relates to computers, input/output devices, storage devices, and technology for information transmission.

In discussing hardware, it should be noted that the idea of the computer as a discrete electronic device is out of date. There is a convergence of information technologies to form systems of which computers are components. Two terms have been coined to characterize this integration:

télématique--combination of computers and telecommunications (Bright/52, Mirabail/335)

videomatics--interconnection between computers and audio/video technologies such as television (Sebestyen/437).

A project seeking to promote these linkages is the Inter-institutional Integrated Services Information System (INSIS) project of the European Communities/147. At the moment telephone, telex, and data transmission take place on separate networks, using incompatible equipment. Advances in technology mean that an integrated network is now possible. The CEC plans to incorporate existing telephone, telex, data transmission and teleconferencing and some new services such as electronic mail, teletext, videotex and videoteleconferencing into a new network. The objectives of this project are to harness new information technologies to improve the transfer of information between Community institutions and member state governments, to improve productivity, and to stimulate development and use of these technologies through creation of a large test market. It is hoped that this project will provide a spur to adoption of common standards for new technologies and avoid incompatible national standards. The project was being planned during 1982 with development and implementation scheduled for 1983-86. In 1974 Penniman et al/385 suggested the need to distinguish among computer, communication and information networks as a function of the units which were interconnected and the type of links. As the description of INSIS suggests, quite complex networks, not easily classifiable, are evolving. The developments making this possible are discussed below in the sections on networks and machine-machine interfaces.

Compatibility of computer hardware refers to the ease with which a computer program running on one machine may be made to run on another machine. Hardware compatibility is achieved through similarity of instruction sets (or the ability to simulate similarity of instruction sets) (Meek/324). There are two ways to achieve hardware compatibility: (1) use identical instruction sets in both machines; (2) use special instructions in the second machine so it recognizes instructions which it does not have (or which it does have but will not produce the same result when executed) in the code being transferred. These instructions are then translated into instructions on the second machine which produce results identical to those of the first machine. "Upward" compatibility refers to the amount of similarity of a newer or bigger computer compared to an older or smaller one and relates to the possibility of transferring programs from one machine to the other. Similarly "downward" compatibility refers to the transfer of programs from the newer (bigger) computer to the older (smaller) one.

With respect to achieving hardware compatibility, competitive practices of various manufacturers may be counterproductive. While computer manufacturers may seek to achieve compatibility within their own lines of equipment, they may make an effort to minimize compatibility with their competitors' equipment. A similar situation can exist for other hardware components such as peripherals and storage devices. In building up a network of devices, there are two ways to achieve compatibility: either compatible devices must be purchased in the first place or machine-machine interfaces must be developed which would allow the devices to communicate (Woods/559).

5.1 Computers

Recent developments in computer hardware have led to the availability of three families of computers: mainframes, minicomputers, microcomputers. Although the classification presented below suggests that it is possible to clearly differentiate microcomputers from minicomputers from mainframes, rapid developments in technology are blurring these distinctions. The basic characteristics have been outlined by Matthews/316:

	Micro	Mini	Mainframe
number of terminals			
1-4	x		
4-80		x	
80+			x
number of simultaneous tasks	one	limited	large
hardware capabilities			
CPU word size in bits	8-16	16-32	32-64
size of CPU memory	limited	moderate	large
I/O data speed	slow	medium	fast
type of processing	slow serial	medium overlapped	fast parallel
auxiliary storage device	floppy diskettes	small-med. hard disk	large hard disk
type of applications			
low cost computing	x	x	
transaction processing		x	
major computing			x
batch processing			x
speed of application			
slow	x		
fast		x	x

These characteristics affect the types of applications which computers in each family can support. 16-bit microcomputers are becoming more common (Lu/[294]) and hard disks are now available as auxiliary storage for microcomputers. The CPU is the part of the computer controlling interpretation and execution of processing instructions. In microcomputers with an 8-bit word size, the instruction set is designed such that the basic unit of information manipulated by the computer is 8 bits in length. Among 8-bit microcomputers different architectures provide different instruction sets and thus can lead to incompatibility among microcomputers from different manufacturers. Now that 16-bit and even 32-bit word sizes are becoming available, microcomputers can achieve compatibility with minicomputers and even mainframes in the sense that software written for the large machines will also run on the microcomputer (Fosdick/[160]). However, such software must not make demands on the computing environment which exceed the capabilities of a microcomputer, e.g., the limitations on memory size and input/output data speed. Large machines usually have special appendages to keep the rate of flow of information into and out of the machine commensurate with its ability to process information.

Compatibility in computer hardware may be achieved either through standardization or adherence to de facto standards. Standardization as a basis for hardware development is illustrated in the approach adopted by six CMEA (Council for Mutual Economic Assistance) countries--Bulgaria, Czechoslovakia, German Democratic Republic, Hungary, Poland, USSR--in the development of a series of upward-compatible third generation computers known as the Unified System or Ryad. Davis and Goodman/[106] estimated that by 1980 the Unified System would be second only to IBM 360/370 in the number of installed mainframes. IBM's dominance of the market has led other manufacturers to adopt its hardware as a de facto standard with the introduction of plug-compatible mainframes and peripheral devices (Bartholomew and Gebremedhin/[28]). Such equipment may be interchanged with a competitor's equipment without requiring modifications to the rest of the system, but it is marketed at a lower price and/or has some enhanced capabilities which make it attractive to the customer. Similarly a plug-compatible peripheral (e.g., magnetic tape and disk drives) is one that can replace an existing unit made by either the mainframe manufacturer or by another independent maker without requiring any changes to either systems or application software (Muldowney/[344]). McGuire/[319] notes the advantage of using plug-compatible machines as remote processors for distributed processing, because computers compatible with the central mainframe and its software can be acquired at a reduced cost.

The growth in popularity of plug-compatible equipment reflects users' reluctance to abandon software developed at considerable cost in exchange for hardware with new features. Birnbaum/[42] calls this the "albatross of compatibility with the past" (p. 764) and suggests that the greatest problem facing both computer manufacturers and users is graceful migration from current systems to newer computer architectures without sacrificing the huge investments in application software. Salton and McGill/[429] describe the development of several types of special purpose search equipment well suited to the processing of large information files, such as associative processors and array processors. The potential of such new designs will not be realized, however, until the necessary hardware and software becomes more widely available at reasonable prices.

Asserting that for the last 30 years the principles of computer system design have remained largely static, the Japanese have begun a programme of research and development for what they term "fifth-generation computer systems" designed to be particularly well suited for the processing of nonnumeric information (Yasaki/[562]). As described by Treleaven and Lima/[488]:

> The planned fifth-generation computer systems may be looked upon as forming a new computer family in which members provide powerful facilities for problem-solving and inference, knowledge-base management, and intelligent input/output. Interfaces are to exist at both the software and hardware levels, allowing software modules and hardware computers to be configured for a specific application or set of applications. The resulting systems will be analogous to Lego building blocks, with each block performing some function and a configuration of blocks functioning as a specialized computer system. (Page 80)

Begun in 1982, the objective of the work is to have a prototype by 1990. Rather than being a "number cruncher" or "data processor", such a system is designed to support information processing through provision of an intelligent interface, knowledge-base management, and problem solving and inference. Although this approach is a definite break with past designs of computer hardware, such machines may be adopted in spite of the lack of "upward compatibility" because of the special features that would be available if the Japanese succeed in meeting their design objectives.

5.2 Input/output devices

In computer systems, input is the translation of incoming information into electrical patterns suitable for computer processing; output is the reverse process in which electrical patterns are translated into a form readable by human beings (Necas/354). This translation process is carried out by the input/output devices of a computer system. These devices may be classified according to the means used for data entry and display (Turtle et al/492):

Input Devices

Tactile
- Keyboards
- Touch-sensitive displays

Optical
- Bar code
- Optical character recognition
- Light pens

Other
- Speech
- Devices for physically impaired

Output Devices

Hard-copy
- Standard paper
- Electrostatic
- Electrosensitive
- Thermal

Transient-image
- CRT
  - Standard
  - Graphics
  - Viewdata
- Flat panel
  - Plasma
  - Light emitting diode
  - Liquid crystal display
  - Electroluminescent

Other
- Audio
- Devices for physically impaired

This section considers how some of the more widely used input and output devices affect information system compatibility.

As terminals are the input/output devices in most common use, it is appropriate to consider them first. Basic functions of a terminal are data entry, data display, and data communication. Although touch sensitive displays are one means of data entry, keyboards are much more common. Necas/355 has described three ways in which keyboards can vary: (1) the arrangement of keys for alphabetic and numeric characters; (2) the arrangement of keys for special symbols and functions; and (3) number of keys. He notes that variations arise due to the varying requirements of new applications, the need for special symbols for certain national alphabets, and the lack of co-ordination among standards bodies. The keys determine the character set available for data entry from a particular terminal. Striking a key causes the code associated with that key to be transmitted; shift keys allow more than one character to be associated with a particular key, such as upper and lower case for alphabetic characters. Thus keyboards can affect compatibility in at least two ways: (1) if keyboards vary, the human user must adapt by learning the new arrangement and content of keys; and (2) keyboards with restricted key sets may not be capable of generating all the characters required for interacting with a particular remote host computer.

In addition to having variations in keyboard layout, terminals can differ in line length, character spacing, print speed, and printing technique. CRTs can vary in screen size, display area, character set, and character format. To be used for a given application, terminals must be capable of displaying data at the appropriate speed (e.g., 300 or 1200 baud) and in the appropriate format (e.g., sufficient line length and number of lines per screen). With respect to information display, Ehlers/137 advocates "user compatibility" in relation to character design and size. Legibility considerations should influence choice of character height and width, the use of lower vs. upper case, and line width. An additional variable is the presence or absence of local "intelligence" and, if present, its capacity (Stewart/465). The intelligent terminal includes a microprocessor and some memory and is capable of being programmed (Tedd/476). Tyner/493 provides a checklist of factors to consider in acquiring print terminals and Crawford/95 provides a similar list for CRT terminals.

Vanderheiden/505 reminds us of groups of individuals not well served by standardization of terminal devices, i.e., those individuals with visual, physical, and/or cognitive handicaps. They may be unable to manipulate standard keyboards or to deal with standard displays of information. He suggests a number of modifications possible to make terminals accessible to the handicapped.

Optical means of data entry include bar codes and optical character recognition (OCR). Different schemes for bar coding exist, each supporting different character sets. For successful entry of bar coded data such as book labels and patron identifiers in a library circulation system, the reader must be able to interpret all the characters used in the bar code scheme. Standards for optical character recognition cover such factors as fonts, materials, and character positioning. Groenewegen and Marshall/190 described a test of OCR as a method of data capture for decentralized information systems such as INIS and AGRIS in which a major proportion of input is prepared in developing countries. IBM Selectric typewriters were used for preparation of data input in a standard font and it was found to be an appropriate substitute in countries that do not have facilities for preparation of magnetic tape input.

Although standardized fonts can be used if data are being keyboarded for the first time, such systems for OCR would be costly for retrospective conversion of data. Within the last few years a much more flexible omnifont scanning device capable of recognizing and flagging font changes has become available. It is the

Kurzweil Data Entry Machine (KDEM) which can successfully read typed or printed materials with a mixture of fonts (Derfall/111). Jennings et al/255 describe the use of this device to capture items from printed editions of Water Pollution Abstracts in machine-readable form. The equipment works directly from a printed page requiring only a clearcut left-hand margin as a start. After a period during which the machine is trained to recognize variants in a particular printed text, the scan becomes automatic and proceeds at a rate at least four times faster than material could be rekeyboarded. Jennings et al found the accuracy of the device for their application to be 99.8%. The KDEM illustrates how advances in technology, in this case pattern recognition techniques, reduce the restrictions needed to achieve compatibility. Instead of being limited to certain standard fonts, texts in any font used consistently within a given text can be processed.

Just as OCR devices are capable of processing a greater variety of printed input, devices for speech recognition are gradually becoming more versatile. It is possible to classify machines into three broad categories defined by the type of spoken material they are designed to accept (Underwood/494): (1) isolated word recognizers require the user to segment his speech; (2) connected speech recognizers allow the user to speak in a more natural manner and enable faster data input but still do not permit transcription of a wide range of normal human speech; (3) speech understanding systems try to mimic human performance but at present fall short in terms of vocabulary size, error rate, and speed. Speech recognition is a difficult task because many factors affect speech, e.g., the position of a word within a sentence, position of a sound within a word, speed of talking, dialect differences, vocabulary characteristics, temporal effects, and extraneous noise. Because there is insufficient information in the speech signal alone to enable decoding, it is necessary to provide the machine with knowledge about speech (phonetic and phonological rules), the language being spoken (grammar and meaning of words), as well as real world knowledge about the topic of conversation. At present the most successful systems are isolated word recognizers and even they are often speaker-adaptive, requiring "training" to recognize the characteristics of speech of the various individuals who will use them. More development is required to produce speaker-independent systems. For the information system user further progress in speech recognition promises to offer alternatives to keyboard input in the future.

5.3 Information storage devices

Devices or media for information storage include punched cards, paper tape, magnetic disks, magnetic tape, microforms, and videodiscs. Not surprisingly, compatibility is more of a problem for the videodisc, developed quite recently, than for the other media for which standards are available. Compatibility of information storage devices is of particular concern when they are used as a basis for transferring information. If data transmission via telecommunications links is not possible, then information storage devices must serve as intermediate carriers of information. Examples of this approach are the machine-readable output on tape provided to Bell Laboratories by an online vendor as an alternative to printouts or direct transmission (Hawkins/201) and the planned BITS (BIOSIS Information Transfer System) service of Biosciences Information Service which will distribute selected citations from the BIOSIS data base on floppy disks to individuals owning microcomputers with appropriate retrieval software. Because magnetic tapes and disks are widely used in information systems, this section considers characteristics which affect their compatibility. In addition the characteristics of microforms and their potential rival videodiscs are reviewed.

Magnetic tape is widely used as a low cost, large capacity information storage medium. Wilmot/552 has identified several tape standards and analyzes the extent to which a number of bibliographic data base producers comply with these standards. She found that the most common physical format was ½ inch wide tape with 9 tracks and a recording density of 800 rows per inch with odd parity for error checking (i.e., a parity bit in each character code is set either to zero or one in order that each row contains an odd number of ones). In this the data base suppliers comply with ISO standard 1863 for 9 track, 12.7 mm (.5 inch) wide magnetic tape for information interchange recorded at 32 rpmm (800 rpi). Any organization with a tape drive accommodating the standard could then read these tapes. Difficulties arise with variations in the number of tracks (7 rather than 9) and in the recording density (e.g., 800 or 1600 rpi). Because tape drives can only read/write certain types of physical tape format (number of tracks and recording density), drives which can accommodate these variations must be available or the tape must be rewritten in a standard form.

Additional variations arise because records can be grouped in different ways on the tape to form blocks, which are handled as units in reading the tape. Records may be unblocked (one record = one block), blocked (at least two records form a block), or spanned (one record occupies more than one block). In addition blocks can vary in size. The maximum block size found in a file of records must be known so that the software can reserve sufficient space in the CPU. In addition, variations in programs to process records must be written if different tape suppliers block records differently. Wilmot notes that differences in blocking characteristics may exist for practical and economic reasons, such as the type and make of computer and the size and number of bibliographic records contained on tapes. A conversion program may be required to change the blocking characteristics to meet local processing needs. For example, it is probably inefficient for machines with large core memory to process tapes with short unblocked bibliographic records. Conversely, computers with small core memory may not be able to handle large unblocked records.

In addition to Wilmot, other people interested in processing tapes from multiple sources have described the diversity of tape formats frequently found. Schipma et al/[433] analyzed seven machine-readable bibliographic data bases and found variations in blocking factors and block size. Levine/[286] notes that each data base to be put up by an online vendor must be analyzed individually and tailored programs written to handle its unique format. As Kent/[261] observes

> The essentially moronic nature of computers makes it necessary to define exactly the form in which information is recorded on tape and to ensure that the actual records on the tape conform to these formats. (Page 505)

Just as tapes have proved to be a useful medium for information exchange among mainframes and minicomputers, floppy disks support exchange of information among microcomputers. Floppy disks come in two standard sizes--5¼ inch and 8 inch. In order to read data recorded on the disk, a drive of the appropriate size must be available. Floppy disks may vary in density and have data stored on one or both sides. Variations in format of data storage can lead to the inability to read data from a floppy with data recorded by a different microcomputer. Davis and Rush/[105] provide illustrations of storage of data on tape and disk and the variations in format which can lead to incompatibility.

McMurdo/[321] contrasts the characteristics of magnetic media and microform. Magnetic media are speedily accessible, readily updateable, and capable of extensive manipulation. However, such media are expensive and limited in capacity. Microforms, on the other hand, are inexpensive, virtually unlimited, and superior for diagrams, illustrations and photographs which are not easily digitized. However, microforms are a static medium, not readily updated. As each medium has advantages and disadvantages, it is likely that the most effective systems will be integrated ones that make use of a combination of storage media (Neary/353). Computer-output microform (COM) represents a linking of these forms in which data are transferred directly to microform from a tape rather than being photographed from printed pages. With respect to the possibility of word processor output microform, Merali/328 notes that the lack of standardization of floppy disk formats in the word processing industry and the slowness of floppies have deterred COM service providers from using them.

As the micrographics industry has matured, early tendencies toward proprietary design have given way to standardization (Saffady/423). Neale/352 provides a bibliography of standards approved by a variety of organizations and agencies, covering such things as film dimensions, reduction ratios, and composition. Adherence to such standards insures compatibility and interchangeability of supplies and components. In contrast to microfiche and roll film, lack of standardization has deterred development of cartridge and cassette applications.

Unlike microforms, videodiscs are a large volume information storage technology lacking standards, with the result that there are a number of incompatible systems on the market. Suthasinekul/470 provides a comparison of the properties of microform and videodiscs as storage media using a variety of parameters such as access/retrieval time, storage compaction, and life expectancy. The videodisc is an attractive medium because it features high information storage capacity, low information storage cost, and rapid random access (Barrett/26). In addition, the data stored have a long lifetime (Hecht/204).

Sebestyen/437 presents a classification useful for comparing videodisc technologies: (1) information storage may be analogue or digital; (2) information pickup may be mechanical (contact) or nonmechanical (noncontact); (3) information recording may be accomplished by mechanical replication or by laser beam recording. Whereas digital encoding is suitable for encoded alpha-numeric material (text, data and computer programs) and facsimile-type static pictures, analogue encoding is suitable for free-frame pictures, motion pictures, and audio information. Tressel and Penniman/489 suggest that a medium which can support digital data storage will have a bright future because of the trend toward digital transmission. Heath/202 compares the characteristics of currently available systems. Four incompatible systems that use two completely different technologies are competing for market acceptance: capacitance groove, capacitance nongroove, reflective laser, and transmissive laser. The key distinction between laser and capacitance systems is the method of reading information. Laser systems are noncontact devices whereas capacitance systems use a stylus. In laser-based systems reflective devices read one side of the disc at a time while transmissive devices allow the laser to read both sides without manually turning over the disc (Paris and Boss/381). Development of videodisc standards is needed to insure interchangeabilty of data stored on discs. The result of an overabundance of incompatible systems is to hinder public acceptance of videodiscs as a storage medium (Edelhart/135).

Goldstein/181 describes an intelligent videodisc interface unit developed by the Lister Hill National Center for Biomedical Communications at the National Library of Medicine. It allows computerized control of videodisc players via a set of higher level commands. A degree of device independence is achieved by allowing different types of players to be interfaced without changing the higher level control commands used by the computer program.

Marsh/311 has suggested that in using videodiscs for information retrieval, inverted files could be stored on magnetic disks and the data to be retrieved on videodiscs. In this way, data in the inverted files, including indexes which are subject to change, could be modified by regenerating information on erasable magnetic disks while searchable data would be permanently stored on videodiscs. An example of this approach is the Video PATSEARCH system which provides access to 750,000 United States patents issued since 1971 and consists of a set of videodiscs, a terminal, a text printer, an interactive videodisc player, and a telecommunication interface (Schulman/434). Options include an auxiliary display monitor to permit display of text and drawings simultaneously and a video printer to obtain copies of drawings and chemical structures. Text information is stored at a remote host computer and graphic images are stored on videodiscs. The system is designed to provide rapid, interactive retrieval of patent documents using any of the information on the front page, including keywords from abstract and title, company names, inventor names, classification codes, or other text supplemented by random access to the associated drawings on videodisc. A videodisc interface unit plus microcomputer-based display terminal provide for automatic selection and display of drawings associated with retrieved records. The patent number is linked to the location of the appropriate drawing through a mapping available on the microcomputer floppy disk. The software controlling the disc player, terminal displays and mapping program has been written in Pascal to reduce hardware dependence.

5.4 Character sets

In the context of information systems, a character set is the collection of numeric, alphabetic, punctuation, and special symbols used by an information system (Dippel and House/121). Each character has two forms: a human-intelligible graphic form and a computer-intelligible form consisting of a group of binary bits. With respect to a particular piece of equipment, the character set is the collection of characters available for display or processing on a computer or peripheral device. In addition each programming language has its own character set which defines the set of characters used in writing programs in that language (Lee and Ralston/281). The character set for a programming language is not necessarily the same as the character set allowed for data; the latter can be much larger. The implementation of a language in a particular computer may not allow certain characters or may allow characters in addition to officially defined ones. Because a given string of bits could be interpreted by a computer as either a number or as the code for a character, most programming languages provide a means of indicating the data type of each variable so that when the contents of the variable are accessed, the computer interprets it in a proper manner (Shneiderman /443).

Although a 6-bit code would allow for the representation of 64 characters within the machine, 7 or 8 bits are required for both upper and lower case, numerals, punctuation, and other special symbols such as diacritical marks. In addition the character set must provide representations for certain control characters which do not print or display but which control the operation of devices (e.g., line feed, carriage return). Two widely used codes are ASCII (American Standard Code for Information Interchange) and EBCDIC (Extended Binary Coded Decimal Interchange Code) (Flores/155). As its name implies, ASCII is a standard 7-bit code designed for exchange of data between systems and it has been adopted by ISO as ISO 646, 7-bit coded character set for information processing interchange. ASCII-7 can be modified to produce an 8-bit code known as ASCII-8 in which bit position 8 is zero. EBCDIC is an 8-bit code developed by IBM. One difference between the two is that the collating sequence for EBCDIC has numerals following letters, while for ASCII the reverse is true. Programs do exist for converting EBCDIC to ASCII and vice versa. In order to remain compatible with IBM, many computers use EBCDIC for internal processing. Whether ASCII or EBCDIC is used, an additional parity bit is added when data are encoded on 9-track magnetic tape as a means for detecting errors. Although ASCII is usually used for transmission of data on tape, most actual processing uses EBCDIC. However, only 96 characters are standard in EBCDIC; additional characters may be represented in different ways as described by Crawford/96 in his comparison of codes used in processing by OCLC, the Research Libraries Information Network, and the Washington Library Network. He notes that, while it would be difficult to standardize EBCDIC because variant usages are embedded in existing software, in each case one can do an EBCDIC/ASCII-8 translation to create transportable data files. He advises that any effort to transmit bibliographic records in EBCDIC must be accompanied by full documentation.

Character sets developed for conventional data processing applications do not suffice for libraries (Kohl/271). The Roman alphabet used for English must be augmented by accents and diacritical marks to represent the extended character sets used by other languages. When a character with a diacritic is encoded, the code for the diacritic can either precede or follow the code for the letter upon which the diacritic is to be superimposed for display purposes. Although proposed standards for videotex suggest that the code for a diacritic should precede the code for the letter, Sherwood/441 argues that diacritics should come after the letter. He justifies his argument by pointing to a number of disadvantages to the prefix approach in a computer environment: inappropriate alphabetization when using simple sort algorithms, unnaturalness in typing diacritics, and extra complications in handling automatic raising of diacritics on upper case letters. Given his experience handling many alphabets on the PLATO system for computer-assisted instruction, Sherwood notes that these disadvantages would not be found if the code for diacritics were entered immediately following that for the letter. For display an automatic backspace would place the postfix diacritic on a letter, with automatic raising on upper case letters.

Even more difficulties are introduced when attempts are made to handle non-Roman alphabets. One way of extending a limited character set is by means of escape characters: a non-printed character that causes the subsequent characters to be interpreted differently. For example, escape characters are used to accommodate non-Roman alphabetic character sets in the MARC communications format/237. Presence of escape characters indicates that codes that follow will represent letters in a specific different alphabet instead of the Roman letters that the codes would otherwise stand for. Thus Roman, Cyrillic, and other alphabets with approximately the same number of characters may reuse the same patterns of code. ISO 2022 provides a standard for the creation of extended character sets and ISO 2375 specifies a procedure for registration of escape sequences.

The need for escape characters arises when multiple alphabets are to be used in the same system. Thus the American approach to the Hebrew alphabet (which has 27 characters), for example, requires escape characters to shift from Roman to Hebrew and back again as needed. In contrast the Israeli data processing community has adopted the simpler solution of incorporating Hebrew characters within existing character sets, usually instead of lower case Roman (Adler/1). This pragmatic arrangement allows display of both upper case Roman and Hebrew characters with no need for escape characters and minimal hardware adjustments. CRT terminals and matrix printers need only be supplied with alternate ROM (Read Only Memory) to display Hebrew in place of lower case Roman.

Although an 8-bit code incorporating the use of escape characters can accommodate multiple alphabets, it is an inadequate solution for the thousands of ideographic characters used by East Asian writing systems. Longer codes are required to represent each character uniquely (Agenbroad/2). Morita/339 provides a review of approaches to Japanese character input describing characteristics of the Japanese language that must be dealt with. He notes that each Kanji (Chinese ideograph) is represented by a two-byte (16-bit) binary code rather than the one-byte (8 bits) found in ASCII. Given the choice of a two-byte code, problems to be solved include which Kanji to represent, what code to assign, and how to represent Kanji on a keyboard. In the early stages of automation in Japan, different organizations developed independent systems for encoding Kanji which were mutually incompatible. Standardization was necessary to support exchange of information among systems, so the Japanese Industrial Standard C6226-1978 for selection of characters and assignment of codes has been developed. Institutions with different codes must translate their own internal schemes into the standard. In addition to codes for 6349 Kanji, the standard specifies codes for 169 Kana (Japanese straightlined syllabary for borrowed words and names). A given information system may choose to implement only a subset of the standard. For example, the Japan Information Center of Science and Technology (JICST) has selected 158 of 169 Kana and 1879 of 6349 Kanji based on their frequency of occurrence in the scientific and technical literature (Takano/473). Wang and Dunn/523 provide a bibliographic survey of Chinese character computer encoding methods. Although several countries are working on the problem, a single standard has not yet emerged.

Choice of character set has an important effect on all equipment used to generate or print text. While having codes to represent various character sets is important for transmission of bibliographic data, it is also desirable to have input/output devices capable of handling both an extended Roman alphabet and other alphabets such as Cyrillic and Hebrew. A terminal recently developed specifically for the library market is the Telex 476L which can accommodate extended Roman as well as Cyrillic and Hebrew, with provision for diacritics, subscripts, superscripts, and special graphics to accurately display entire bibliographic records /539. Diacritics are displayed as part of the associated characters instead of beside characters as frequently found on other terminals. Other organizations such as the Research Libraries Information Network in the United States are working on the more difficult problem of building a terminal to handle Chinese, Japanese, and Korean scripts. Miller/334 suggests that, where input is in the original script (e.g., Cyrillic), computer programs could contain the various romanization schemes for scripts and create romanized forms automatically when processing data to allow interfiling for such applications as indexing and authority control.

For meaningful interchange of texts, the sender and receiver must agree about character codes. The existence of character sets differing in the characters included and in the codes used to represent them presents an impediment to more extensive communication of texts. Two recent studies propose a "Unified Coding Scheme" (Ross/417) and "un code universel"/145 to deal with this problem. As Ross notes, text communication is possible if one of three conditions holds. The character sets can be: (1) identical--the same in every respect; (2) compatible--one is a subset of the other; (3) interworking--communication possible between two character sets that have equivalent or subset repertoires of graphics and controls but coded differently; conversion rules can be implemented. Ross notes that, with the advent of microprocessors built into various pieces of equipment, interworking becomes a viable alternative to compatibility or identity. He identifies four basic approaches to standardization of character codes:

(1) subset of characters common to all national requirements is standardized and other characters needed within each country are left undefined;

(2) subset that is common to all national requirements is specified together with some means of regulating additional characters needed by each country;

(3) comprehensive set of characters is defined to serve all national requirements and to be used within each country in its entirety for international communication;

(4) comprehensive coding scheme is devised in which the defined character set comprises a superset of all national requirements and specific proposals are made to introduce relaxations to suit national requirements and implementations on particular types of equipment.

Ross advocates the fourth approach as a means to accommodate varying national needs and the limitations of various applications equipment such as terminals for videotex and teletext as well as printers and CRTs. A coding interface would convert from local usage to the comprehensive coding scheme where necessary. The study by CREL-France/145 suggests a "Euroalphabet" along the lines of solution 3 above, sufficient to meet the needs of the European community by supporting multilingual interchanges. In addition to standardizing codes for characters, the study recommends standard codes for languages, countries, currencies, units of measure and mathematical symbols.

5.5 Information transmission

Familiar technologies for information transmission--the telephone, telegraph, and television--are now supplemented by networks for digital data transmission, facsimile, videotex, and teletex. Although these new technologies offer a variety of new capabilities and promise to support new types of information services, they also present problems due to the incompatibility of nonstandard equipment. After an overview of alternative means for data transmission, this section considers the characteristics of the new technologies in more detail.

In a communications network the characteristics of the links or channels govern the way in which information is transmitted between nodes. Possible channels include wire pairs, coaxial cables, optical fibres, microwave links, and satellites. These vary in bandwidth and thus in both capacity and speed of transmission. Optical fibres in particular have a high capacity. However, fibre optics has developed so quickly that standards are only now being prepared (Hecht/203). For example, standards development in the United States is being pursued by the Electronic Industries Association for commercial fibre optics and by the Department of Defense for military systems.

In most cases, a satellite system does not provide any facility that cannot conceptually be provided by more conventional channels. The satellite finds application in situations where it is a more cost effective solution than the alternatives, e.g., transocean routes, domestic systems over difficult terrain or large areas, and systems in developing countries where there is no previous or little existing communications infrastructure. Satellite communications now represent a proven technology, and system configurations with small, inexpensive earth terminals for the reception and transmission of electronic messages can provide new opportunities for a variety [illegible] users and uses. As a recent review of the use of satellite communication for information transfer points out, the main problems relating to the use of satellite services for the transfer of information do not fall in the technical field (International Institute of Communications /240). Instead they concern such areas as the organization of the services, the integration into overall telecommunication policy, the development of telecommunications regulation, and financing and costs to the user.

A signal transmitted may be analogue (continuous) or digital (discrete). The latter is compatible with computers and terminals. For analogue transmission a modem must convert from digital to analogue when data are entered and from analogue to digital when the data reach their destination. The mode of transmission can be simplex (message travels in one direction only), half duplex (transmission in either direction but not simultaneous), or full duplex (transmission can take place in both directions simultaneously). In asynchronous transmission characters are sent one at a time and every character has a special element at beginning and end; in synchronous transmission signals are sent at precisely synchronized time intervals. Kenney/260 distinguishes line or circuit switching, where the communication path is established before a message is accepted for transmission, from message switching, where a message is accepted, stored, and sent on as a channel becomes available. Packet switching is one type of message switching in which the message is broken into packets with information about their origin and destination. At the destination, packets are reassembled. Packets have a fixed maximum size in a given network, e.g., 1008 bits on Telenet and 2040 bits on Canada's Datapac (Thoma/482).

Rigg/408 has reviewed the characteristics of various types of networks from the point of view of their usefulness for data transmission. The public switched telephone network, though well suited for voice communication, has a number of disadvantages with respect to data communication: very high error rates, inability to support high speed terminals, cost based on distance and call duration rather than on the amount of data transmitted. In contrast packet switching is lower cost, has fewer errors, and a higher transmission speed. In addition it has facilities for speed changing and procedural conversion, thus allowing communication between different types of terminals and interconnection of networks (Raitt/401). Brownrigg and Lynch/59 have described the development of a private, packet-switched network to support terminal access to the machine-readable Union Catalog of the nine campuses of the University of California.

A protocol is the set of rules governing exchange of information between two systems. Protocols must be adhered to by both sender and receiver. They govern such things as mode, speed, character length and code which affect the format and timing of data transmission. The protocol for communication between two computers may differ from that between a computer and a "dumb" terminal. Brooks/54 notes that some existing online systems are not set up to handle "smart" receivers and will not respond to user checks of data transmitted.

Before turning to a discussion of facsimile, videotex, and teletex, the major bodies involved in standardization of data communications at the international level should be identified. These include the Comité Consultatif International Télégraphique et Téléphonique (CCITT), an agency of the International Telecommunication Union, the Conférence Européene des Administrations des Postes et Télécommunications (CEPT), which is the European telecommunications authorities' standardizing body, and the International Organization for Standardization. In the United States both the American National Standards Institute (ANSI) and the Electronic Industries Association (EIA) are also important. Folts and Karp/157 provide a compilation of data communications standards issued by CCITT, ISO, ANSI, EIA, and the United States federal government.

Facsimile is a system for the transmission of images. The image is scanned at the transmitter, reconstructed at the receiving station, and duplicated on some form of paper. An advantage of facsimile is the ability to transmit printed as well as graphic material without different encoding (Voos/519). This means that graphics can be handled as an integral part of the text and non-Roman alphabets can be transmitted. Systems for facsimile transmission have three components: (1) a source terminal (transmitter); (2) a receiver (where communication terminates and the copy is recorded) and (3) the medium through which signals travel. The transmitter and receiver are usually incorporated into one machine, the transceiver, although send only and receive only machines exist. In spite of the existence of standards developed by industrial and governmental organizations at both the national and international level, several of the various makes of commercially marketed facsimile terminal units (transmitters, receivers, transceivers) are incompatible (Costigan/[92]). For example, devices may adhere to the EIA standard for resolution and scan rate of 96 scan lines per inch and 180 lines per minute and still differ in modulation (FM or AM) or in phasing and synchronizing schemes. Barrett/[27] has observed that different machines are unlikely to be able to communicate directly even if using the same modulation techniques because until recently manufacturers seem to have insured that machines can only communicate with those made by the same manufacturer. In addition to analogue devices, digital devices are now available, differing in code formats, data compression schemes, and error detection techniques. Furthermore, there are differences in resolution, scan rate, and direction of scan between the U.S. and European systems.

Although the easiest approach to insuring compatibility in facsimile transmission applications is to have identical machines as transmitters and receivers, mechanisms for conversion have emerged which allow communication between devices which would otherwise be incompatible. McKean/[320] reports that several private telecommunication networks have developed software to overcome incompatibility between facsimile terminals employing different transmission speeds and modulation techniques. Facsimile messages are transmitted to a central computer which converts the message for retransmission to otherwise incompatible facsimile receivers. Such conversion services and devices are viewed as an integral part of the development of the ARTEMIS (Automatic Retrieval of Text from Europe's Multinational Information Service) concept for document digitalization and teletransmission (Norman /[374]). Some will be "black boxes" placed between a terminal and the network; others could be located at network nodes or at hosts serving as sources for document delivery. The concept of ARTEMIS is not limited to facsimile as sources could be digitalized either as text or facsimile and stored in data bases attached to host computers from which they would be retrieved in response to a user's request.

Videotex is a combination of existing technologies: computers, telecommunications, microelectronics, and television display systems (Childs/[77]). Videotex is a generic term which encompasses five types of systems (Criner and Johnson-Hall/[97]):

(1) narrowband broadcast (one way)--information is inserted into the vertical blanking interval of the television signal or subcarrier channels of FM radio signals (frequently called teletext);
(2) broadband systems (one way)--full broadcast channel for transmission using either cable television or over the air;
(3) narrowband interactive (two way)--telephone network is used for information transmission;
(4) broadband interactive (two way)--two way cable television is used;
(5) hybrid systems.

The decoder for videotex signals may be a special terminal, a settop adapter, or built into a television. Teletext differs from interactive videotex in the amount of information which can be accessed: a few hundred pages can be broadcast whereas thousands of pages can be accessed interactively.

Videotex supports transmission of visual information, of which five types can be distinguished (Chitnis and Costa/78):

(1) general text--can be reformatted at display without loss of meaning;
(2) positional text--cannot be reformatted on display, e.g., tabulated data;
(3) mosaic graphics--pictures defined by filling character positions on the screen with mosaic elements;
(4) geometric graphics--pictures defined in terms of geometric primitives transmitted to terminals as drawing commands;
(5) photographs--still photos in black and white or colour.

All current teletext and videotex systems use ASCII for encoding text information. Incompatibilities arise due to the way in which graphics are encoded. Alphamosaic systems can be thought of as first generation systems for which the two competing technologies are represented by the British Prestel system (Childs/77) and the French Antiope system (Marti/312). The French system produces most, but not all, of the Prestel graphics functions and adds some features not found in Prestel (Clarke and Childs/83). Alphageometric systems are second generation systems with enhanced graphic capabilities as introduced by the Canadian Telidon system (Tenne-Sens/478). The alphamosaic character set has 8-bit codes for two tone colouring of squares and a screen of 20 lines by 40 characters is created character by character, line by line, filling the screen from top to bottom. It is limited in the detail with which images can be drawn. In alphageometric systems, on the other hand, much of the computing burden involved in producing graphics is in the receiving terminal (Tenne-Sens/478). The terminal is a microcomputer with input/output ports, read-only memory, random access memory, and a microprocessor. Picture description instructions (PDI) tell the terminal what to draw and where on the screen. Geometric forms include points, lines, rectangles, arcs, and irregular polygons.

Marti and Schwartz/313 describe the difficulties of establishing standards for videotex, in particular because efforts at standardization began after large investments in incompatible technologies had already been made. As Smirle et al/449 have commented:

> The challenge of standardization is to foster the development of a healthy service as early as possible without stifling the innovation so necessary to the ultimate goal of a mature network providing a diverse range of services to as broad a market base as possible. (Page 273)

At present there are two major competing standards seeking acceptance in the marketplace. In North America the Bell system videotex presentation-level protocol has a graphic coding scheme essentially that of Telidon with some extensions in keeping with the Telidon philosophy of extensibility of features while maintaining compatibility with established data bases and user hardware (Tenne-Sens/478). In addition to alphageometric graphics, characteristics of this de facto North American videotex standard include:

(1) Alphamosaic graphics character set which permits access to videotex data encoded in the European format. However, some preprocessing of European data is required since codes for colours, text sizes, etc. differ from European codes;
(2) Dynamically Redefinable Character Set (DRCS) permits the host to define shapes of special characters at the beginning of a session and to download them into the DRCS section of the terminal's decoder;
(3) Macro PDIs can be defined, such that a two or three byte identification code can generate a previously transmitted sequence of PDIs;

(4) A wide range of colours is possible.

The French and British have reached a compromise videotex presentation standard by working through the CEPT, but it differs substantially from the North American standard. The CCITT is examining compromises between the two but, as Neustadt/[365] observes, manufacturers and data base providers on both continents are committing too much money to make further change easy. Development of a single worldwide standard therefore looks unlikely. One possible way to accommodate both would be to have a protocol whereby the videotex computer would query the terminal concerning which standard it uses and then send appropriately formatted information. However, this solution would mean that hosts would have to be able to support conversion among all the coding schemes.

Without compatibility, consumers will have to buy several different terminals to participate in all teletext and videotex services. Chitnis and Costa/[78] advocate an approach to terminal design that would maximize compatibility. They suggest that the visual feature capabilities of videotex should be arranged in a Layered Capability Structure. There would be a hierarchy of five types of terminals: A. general text only; B. A + positional text; C. B + mosaic graphics; D. C + geometric graphics; E. D + photoimagery. Such a hierarchy would maximize what they term forward and upward compatibility. In forward compatibility present day terminals can receive and decode expected future information within the limits of their capability; with backward compatibility future terminals must be able to access old data. In upward compatibility low feature terminals should be able to display as well as possible any information from a sophisticated data base; in downward compatibility a sophisticated terminal should be able to access simple data. As an example of this principle they cite television signal standards which allow both monochrome and colour television receivers to display monochrome and colour television signals. Ciciora/[80] has stated the implications of these two forms of compatibility for the development of decoders: at any given time it should be possible to support a range of decoders offering a variety of features; over time it should be possible to add features to the system in a compatible manner so that future improvements do not cause older decoders to malfunction.

An interesting example of a system merging full text and videotex features in support of information retrieval has been described by Griffith/[189]. Battelle's BASIS data management system was merged with Infomart's Telidon system to provide the flexible keyword search and full text features of BASIS plus the videotex graphics capabilities of Telidon. New capabilities for information retrieval provided by the hybrid include the addition of colour text capabilities for a BASIS session dialogue and high quality coloured graphics output capabilities from BASIS data bases. New capabilities for videotex include the flexibility of a keyword search/command language retrieval in place of the less flexible tree structure search method and the ability to support high volume text data bases which are difficult to support in videotex. The interface was designed using the BASIS dialogue file feature which stored Telidon PDI directives along with regular ASCII data for messages and command keywords.

Yet another new technology for information transmission is teletex (improved telex) for text communication. Gabler/[168] describes the development of this service in Germany where the objective has been to develop an internationally compatible high speed text communication service. In the terminal for teletex there is a clear separation between the local part and the communication part. The local part is like a normal office typewriter with all functions familiar to a typist. The communication part is designed for compatibility on a worldwide basis. The clear interface between local and communication parts allows future developments and modifications on the local side without necessitating changes on the communication side. On the other hand, improvements and changes of the communication part will be possible without affecting the local part. There should be a capability of interworking with the existing telex service, with the necessary conversions effected on the teletex side. The network must provide automatic code conversion (7-bit to 5-bit), rate conversion (2.4 Kbits/sec to 50 bits/sec) and procedure conversion. The prospects for international communication via teletex have not yet been realized, however, because the international standard is not sufficiently precise to insure that terminals conforming to it can communicate. In developing standards for teletex, there has been an effort to keep code and protocol standards the same as those for videotex wherever possible to facilitate interworking of services and the use of a teletex terminal to display videotex and vice versa. There are significant differences between the two types of terminals for which adjustments must be made. A videotex terminal has a 40 character line whereas a teletex terminal has 80 characters per line, so teletex lines must be wrapped into two videotex lines. On the other hand, there are no colour mosaic graphics on teletex, so mosaics would have to be represented by some fallback character on teletex.

5.6 Computer networks

A computer network consists of a collection of hosts (computers) providing service to a user (either an end user or another host computer) and a subnetwork providing communication among host computers, users, or both (Kimbleton and Schneider/[263]). Both Neumann/[364] and Malcom and Promisel/[300] are helpful guides to networking terminology. A recently published book by Ahuja/[3] includes an appendix describing four major computer communication networks and their services (Tymnet, Telenet, Datapac, and Transpac).

The proliferation and growth of data base services have been closely tied to the availability of communication networks. The same network technology that supports decentralized access to centralized data bases can also support decentralized access to decentralized (distributed) data bases (Bunch and Alsberg/[64]). Fosdick/[159] points out the difficulty of distinguishing the term "distributed data processing" (DDP) from networking. He suggests that DDP refers more specifically than networking to the distribution of intelligent processors within a single system, whereas networking allows communication between systems. However, the boundaries between networking and DDP are not sharply drawn. McGlynn/[318] includes a chapter describing the network control architecture and data communications protocols of various commercially-available distributed systems.

The computers which make up a network can be configured in various ways. Anderson and Jensen/[10] present a taxonomy for systems of interconnected computers, including such categories as loop, complete interconnection, global bus, and star. Some of the configurations can be found in local area networks (LANs), networks limited geographically and having a high bandwidth in contrast to long haul networks like the public telephone network which are geographically distributed. Kinnucan /[265] compares the characteristics of a number of LANs including the vendor/developer, configuration (e.g., ring, bus), access method, speed, transmission medium (e.g., optical fibre, coaxial cable), maximum distance, maximum number of stations, and packet size. He explains the dual interface approach found in these networks. The network interface handles interfacing functions common to all stations (electrical connection to the transmission medium, message transmission and reception, channel access protocol enforcement). The network interface has two standard faces: one that faces toward the transmission medium and one toward the station. The station interface has a standard face only to the network interface. The station face varies according to the type of station--computer, printer, disk drive, copying machine. Any device with an interface to the network can communicate with any other device possessing such an interface, so that equipment from many manufacturers can be combined on the same network. A LAN can thus solve the compatibility problem for diverse types of equipment by serving as a standard interface. Although LANs solve the compatibility problem within a network, different LANs are usually incompatible. Saal/[422] states that in almost every case manufacturers of LANs have developed their own hardware and software protocols leading to incompatibility. The exception is the Ethernet specification released in 1980 by Digital Equipment Corporation, Intel, and Xerox Corporation. These organizations are encouraging its adoption by other manufacturers. In a library context a LAN could supply individual stations with word processing, ordering systems, accounting, electronic mail, videotex, and other functions while allowing shared mass storage for circulation and cataloguing (Collier/[90]). Possibilities also exist for interconnecting a LAN with long haul networks (Clark et al/[81]).

The possibility of internetwork connection is an area of growing interest in other types of computer networks as well. A gateway is an exchange in one network which allows access to and from another. The primary function of a gateway is to resolve differences between the two interconnected networks in such a way as to permit meaningful communications between users on each of the networks (Cotton/[93]). In both Germany (Zimmermann/[566]) and Great Britain (Heys/[216], Woolfe/[560]) there is interest in having the videotex systems support gateways to external computers. Horne/[225] describes the operation of the Gateway facility on Prestel. Haber/[194] discusses the possibility of interworking between the existing European videotex systems and Euronet. There are four possibilities for interworking, as illustrated in Figure 8: Euronet as a means of connecting customers to a foreign videotex centre, Euronet as a link between videotex centres and the private computers of information providers, Euronet as a means of enabling the videotex terminal to access DIANE data bases, and Euronet as a means of enabling DIANE users with standard terminals to access a videotex centre. In each case such interworking is technically feasible, but modifications of terminals and/or transmissions between hosts and terminals would be required. Télésystèmes-Questel, for example, have installed a new front end processor to handle Télétel terminals which in several respects is a departure from classic teletype-compatible terminals.

Another example of a gateway is that of the International Institute for Applied Systems Analysis (IIASA), providing IIASA users with online access to remote computer resources, data bases and networks, and also to enable remote users from collaborating institutions to access resources available at or through IIASA (Labadi/[275]). Bakonyi et al/[23] describe the Hungarian experience using the IIASA gateway as a means of establishing computer communications between East and West in support of scientific research.

If computer networks are to realize their full potential, they must not be unduly complex for the end user. Just as single system operating systems were developed to relieve the user of the interface burden, so now network operating systems are being designed and developed to assist the network user. The intent is to provide access to network resources as easily and uniformly as single system operating systems provide access to their own hardware/software facilities. Mamrak/[305] describes the design and development of resource sharing services which could constitute parts of a network operating system. One example of a special network facility is the network access machine (NAM) developed by the National Bureau of Standards (Rosenthal/[415], Rosenthal and Lucas/[416]). The NAM is a minicomputer system that acts as a network access point for a user at a terminal and

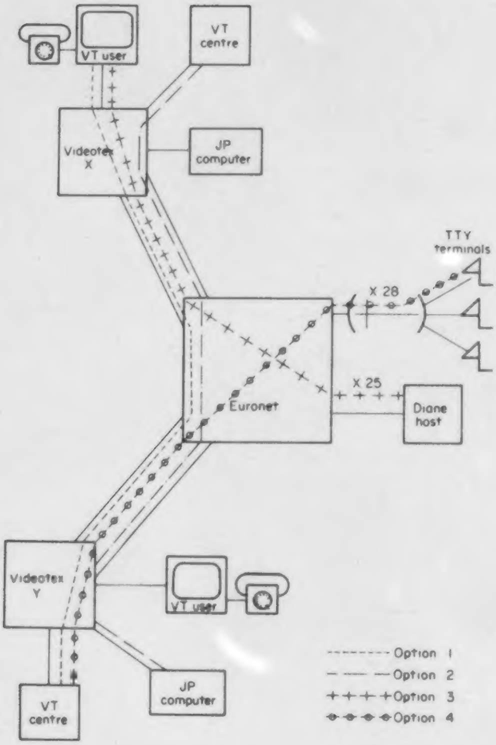

VT user
VT centre
Videotex X
JP computer
TTY terminals
X 28
X 25
Euronet
Diane host
Videotex Y
VT user
VT centre
JP computer
-------- Option 1
— — — Option 2
+ + + + Option 3
⊘ ⊘ ⊘ ⊘ Option 4

Figure 8. Possibilities for interworking: videotex and Euronet. Reproduced from Haber/[194] by permission of IPC Business Press.

assists the user through automatically executing access protocols. Barchanski/[24] suggests provision of standard messages as another network facility to assist the user.

The emergence of universal, shared data networks depended largely on the development of the technology for packet switching (Odegard/[375]). Euronet is one of the better known packet switching networks. Its adoption of the packet mode of operation stemmed from the need to accommodate a wide range of often incompatible terminals on a single network and to provide low cost access in an interactive mode to data bases connected to the network (Kelly/[259]). The packet switching exchanges in Euronet offer two kinds of interfaces: (1) synchronous interfaces for packet-mode data terminal equipment (DTE) to which both host computers and intelligent terminals can be connected; (2) asynchronous interfaces (character-mode) which provide support for teletype-compatible terminals. Euronet supports a series of CCITT protocols. The protocol defines communication between the data terminal equipment (DTE) and data circuit terminating equipment (DCE) (Folts/[156]).

X.25 defines this interface for terminals operating in the packet mode on public data networks. X.3, X.28, and X.29 are a set of protocols defining the connection of asynchronous terminals to the European PAD (packet assembler/disassembler) and the means by which a host computer can access the terminal. A front end processor can be used to handle communication functions for a host computer (Karshmer/258). Purser/399 notes that X.25 has been implemented on front end processors for various machines, in effect a "black box" allowing connection of a host to Euronet. Pouzin/395 asserts that interconnection between two X.25 interfaces is possible only if they are carefully engineered with compatible options because the standard itself provides for incompatible options. An explanation of X.25 can be found in a chapter in Chorafas/79.

Because packet switched networks designed similarly to Euronet are in place or planned for other parts of the world, global internetworking becomes possible. Casey/73 provides a diagram illustrating how these networks could be linked via gateways, as shown in Figure 9. The access procedure developed for Euronet illustrates the level of detail that must be specified to establish the appropriate link between user and host in such a complex network (Mahon/299). The access procedure includes codes for country, network, called device number, number of application on called device, end of field, and the network user identifier.

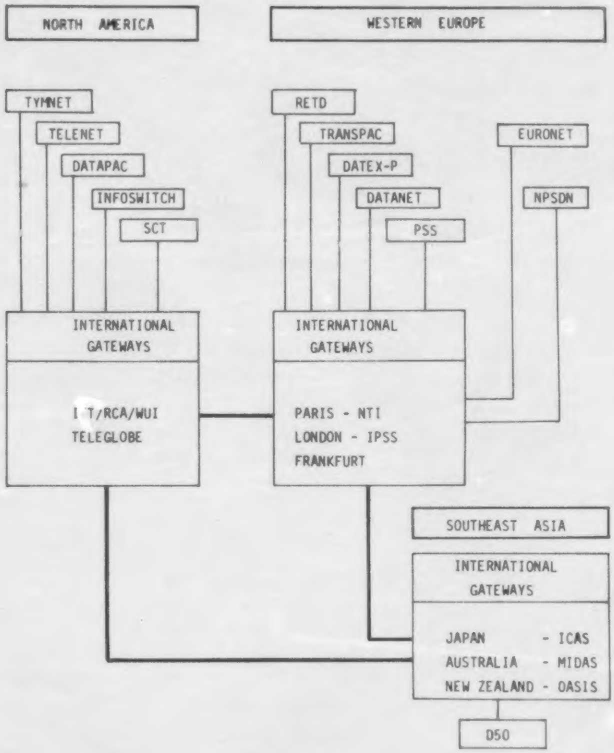

NORTH AMERICA
WESTERN EUROPE
TYMNET
TELENET
DATAPAC
INFOSWITCH
SCT
RETD
TRANSPAC
DATEX-P
DATANET
PSS
EURONET
NPSDN
INTERNATIONAL GATEWAYS
ITT/RCA/WUI
TELEGLOBE
INTERNATIONAL GATEWAYS
PARIS - NTI
LONDON - IPSS
FRANKFURT
SOUTHEAST ASIA
INTERNATIONAL GATEWAYS
JAPAN - ICAS
AUSTRALIA - MIDAS
NEW ZEALAND - OASIS
D50

Figure 9. Global internetworking. Reproduced from Casey/73 by permission of Butterworth Scientific Ltd.

Although packet switched networks are of the same type, in principle one might also seek to use a gateway to switch from one type of network to another, as in the example of videotex to packet switched given earlier (Kirstein/[266]). In this case the gateway may have to perform conversion as well as switching functions (Zimmermann and Naffah/[565]).

5.7 Machine-machine interfaces

Progress made in developing interfaces between machines for other application areas is frequently also relevant to the design of information systems. For example, Eadie/[132] reviews the characteristics of various types of peripheral devices and explains the means of interfacing them with computers. Ailamazyan and Veselovskii/[4] describe the requirements that information exchange via telecommunications imposes on the choice of hardware and software. Whitehead/[541, 542] describes the means of establishing communication links between word processors. Because protocols for these devices are not yet standardized, protocol translators must be used to enable previously incompatible word processing systems to communicate. Hoyt/[231] presents a practical application for the communicating word processor in a library environment.

If interfacing is to be accomplished on other than a situation by situation basis, then some guidelines for standards development must be available. Of particular interest in this regard is the ISO Reference Model for Open Systems Interconnection ISO/TC97/SC16 N227. Open Systems Interconnection refers to standardized procedures for the exchange of data among computers, terminal devices, people, processes, networks, etc., that are "open" to one another for this purpose by virtue of their use of these procedures. Some helpful tutorials describing the concepts associated with Open Systems Interconnection have recently appeared (Buchinski and Islam/[62], desJardins/[112], Gee/[173], Tanenbaum/[474]). The intent of the reference model is to provide a common basis for co-ordination of standards development for each of seven hierarchical layers from which a network is structured.

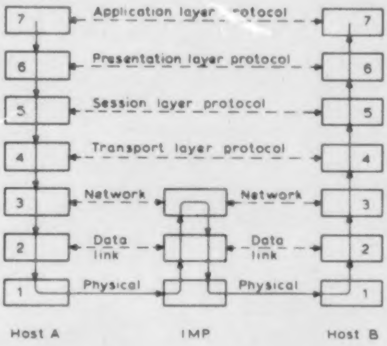


Figure 10. The seven-layer ISO reference model. Reproduced from Tanenbaum/[474]. Copyright 1981, Association for Computing Machinery, Inc., reprinted by permission.

As the figure illustrates, each layer builds upon the previous one. All layers are present on hosts in the network, whereas only layers 1-3 would be found on interface message processors (IMPs) which are minicomputers assisting with communication in long haul networks.

Each layer may be thought of as a program or process (possibly embedded in hardware) that communicates with the corresponding process on another machine according to the protocol governing that layer. In reality data are not transmitted horizontally from machine to machine within a given layer, but instead are passed vertically down layers of the sending machine and up layers of the receiving machine. Only in layer 1 does actual intermachine communication occur. The boundary between adjacent layers is termed the interface. The layers, interfaces, and protocols in the network form the network architecture. A detailed list of the services provided by each layer and the functions in each layer can be found in Appendix B of Meijer and Peeters/[326].

The functions of each protocol can be specified briefly as follows:

1. Physical layer protocol--deals with transmission of the raw bit stream. It defines four sets of characteristics related to a physical interface: mechanical (how two devices are connected), electrical, logical, and procedural (Smith/[450]).

2. Data link layer--converts an unreliable transmission channel into a reliable one for use by the network layer, providing data transmission with error recovery.

3. Network layer--deals with routing and congestion.

4. Transport layer--provides reliable host to host communication for use by the session layer, shielding the customer's portion of the network (layers 5-7) from the carrier's portion (layers 1-3).

5. Session layers--sets up, manages and tears down process to process communications using host to host service provided by the transport layer.

6. Presentation layer--representation and coding of data structures such as characters, display formats, etc.

7. Application layer--depends on users.

Jacobsen/[250] designates level 7 as semantic and level 6 as syntactic to distinguish their respective functions. The ISO Open Systems Interconnection Reference Model says nothing specific about the implementation of these protocols. Instead, it serves as a framework for describing layered networks by specifying each layer--its purpose, the services provided to the next higher layer, and the functions performed. Although it is not a protocol standard, it does suggest places where protocol standards could be developed. In fact many such standards are already in existence. Physical layer protocols include EIA's RS232C (interface between data terminal equipment and data communication equipment employing serial binary data interchange) and CCITT's X21 (general purpose interface between data terminal equipment and data circuit terminating equipment for synchronous operation on public data networks). X.25, mentioned above, represents a network layer protocol. In OSI the layers should not depend on details of a lower layer's implementation. Emmons/[141], for example, describes development of code-independent communication headings at level 3.

Westcott/[538] observes that the advent of small, inexpensive microprocessors allows the introduction of communications protocols into areas formerly closed to protocols. Single logic chips are now controlling the standard layer 1 (physical link) protocols. Microprocessors are performing the work to support some of the more common layer 2 protocols. Furthermore some available microprocessor-based computer systems can support the X.25 layer 3 protocol. The availability of such hardware contributes to the standardization of protocols since the protocol is built-in and is not easily changed.

The lower layers are relatively well understood as exemplified by existing communication protocol standards. Adherence to a common set of protocols allows diverse types of computers to communicate (Newell and Sproull/[367]). The higher levels are presently under study for development of international protocol standards (Bochmann and Gecsei/[44]). Clark/[82] notes that the high level of standards activities is largely due to the rate of innovation associated with the merging information processing and data communications technologies. An important goal of standards programs is compatibility between networks in different countries (Wheeler/[540]). Although it may be possible to convert or map one protocol into another, the availability of standard protocols simplifies the interconnection considerably (Belloni and Le Moli/[35]).

In a tutorial on protocols Pouzin and Zimmermann/[396] have compared the evolution of programming languages and the evolution of protocols. Just as interest in techniques for proving program correctness has developed among those studying programming, there is an interest in verification techniques among those seeking to determine whether equipment which is termed "standard" does in fact adhere to the standard protocol. Both Ansart/[12] and Bartlett and Rayner/[29] suggest possible certification techniques.

Development of standard protocols supports communication or interconnection between components that otherwise would not interact. One environment in which developments have been rapid is the area of office automation. As Spinrad/[461] and Giuliano/177 describe, the electronic desk becomes a professional's link to widely distributed arrays of information sources and services. To succeed there must be representational and protocol standards for creating, displaying, printing, communicating, and filing electronic documents. Tantawi and Mourad/[475] provide design specifications for office work stations including a standardized command language for terminals to simplify their connection to office systems.

Library networking is another area that could benefit from application protocol development. The report on computerized library networking in Canada (Duchesne et al/[128]) recommends that the National Library of Canada concentrate its efforts on development of computer to computer links between shared cataloguing systems. Similarly in the U.S., a Linked Systems Project to develop a Standard Network Interconnection is underway/[57,525]. Funded by the Council on Library Resources, it is a joint effort of the Library of Congress, the Research Libraries Group, and the Washington Library Network. The intent is to develop standard telecommunication protocols for exchange of bibliographic information between separate and different computer systems. Such a system to system link is needed to support exchange of data and/or services between any two or more systems. Avram/[18] suggests that if such a link were standardized, it would be a fundamental building block for library networking in the 1980's just as MARC was in the 1970's.

The general requirements for an application level protocol for bibliographic data exchange were first presented in 1977 (NCLIS/NBS Task Force on Computer Network Protocol/[351]). Until this protocol was developed for use in the bibliographic network environment, there was no means for determining what kind of a message was being sent (Lagueux/276). Little/288 explains the proposed protocol in some detail.

An interface need not always take the form of a telecommunication link. A recent study identifies three possible types of local interfaces to the Australian Bibliographic Network/239. They are a function of the existence of local computer facilities and the degree to which these are integrated into total library management. In the first case, there is no local computer and the library can receive cards or COM catalogues from the network. In the second case, there is a local computer but it does not communicate directly with the network. Output can be sent to it as a magnetic tape. In the third case a local computer is actually connected to the network and transmission can take place directly. Similarly a study conducted by Battelle Columbus Laboratories/30 on the costs and benefits of linking bibliographic utilities (LC, OCLC, RLG, WLN) included direct tape delivery, online "native" mode via dedicated line, and online "translation" mode via dedicated line as the three options. In a "native" mode link a bibliographic utility makes it possible for a member library to use its terminal as though it were part of the other utility's network. The operator must learn the commands and search strategies of the other utility. In "translation" mode the user enters a search request that is converted automatically by the library's utility into standard formats for transfer to the linked utility. Responses from the other utility are then converted by the library's own utility into its own formats. In this link option, the burden of translation is handled by the system. Jones/256 suggests passive links as yet another alternative. Here the user would be required to call up each utility separately for searching rather than having host to host connections established.

In developing a national library network, the technical problems to be overcome include incompatibility of terminals, records, communications lines, line protocols, search languages, commands, search responses, output screen formats, and interactive dialogue between the user and the terminal (Epstein/144). For the purposes of discussion in this report, these various factors have been split into software and hardware elements. But the Reference Model for Open Systems Interconnection suggests a different way of classifying sources of incompatibility and illustrates the dependency of the application level on existence of compatibility (in the form of working protocols) on all the lower levels. It is appropriate to conclude this section with an observation made by Stevens/464:

> While the general compatibility problems between and amongst machines are reasonably tractable and are beginning to yield both to hardware and software developments for more effective convertibility in interfacing and also to standardization efforts, the interfaces between man and machine present both more fundamental and also more difficult challenges. (Page 209)

5.8 Problems relating to hardware

The most notable trend in hardware development is the convergence of information technologies to form systems of which computers are components. Recently coined terms characterize this integration: télématique for the possible combinations of computers and telecommunications and videomatics for the interconnection between computers and audio/video technologies. Quite complex networks of devices, not easily classifiable, are evolving. With respect to achieving hardware compatibility, competitive practices of various manufacturers may be counterproductive. While computer manufacturers are likely to seek to achieve compatibility within their own lines of equipment, they may make an effort to minimize compatibility with their competitors' equipment. A similar situation can exist for other hardware components such as storage devices. In building up a network of devices, there are two ways to achieve compatibility: either compatible devices must be purchased in the first place or machine-machine interfaces must be developed to allow otherwise incompatible devices to communicate.

Compatibility in computer hardware may be achieved either through standardization or through adherence to de facto standards. In the latter case "plug compatible" equipment may be interchanged with a competitor's equipment without requiring modifications to the rest of the system. A major problem facing both computer manufacturers and users is graceful migration from current systems to newer computer architectures without sacrificing the huge investments in applications software. In general, users are reluctant to abandon software developed at considerable cost in exchange for hardware with new features.

Terminals are the input/output devices in most common use. Terminal keyboards can affect compatibility in at least two ways: (1) if keyboards vary, the human user must adapt by learning the new arrangement and content of keys; and (2) keyboards with restricted key sets may not be capable of generating all the characters required for interacting with a particular remote host computer. To be used for a given application, terminals must be capable of displaying data at the appropriate speed and in the appropriate format. Recent developments in optical character recognition (OCR) as a means of data entry illustrate the way in which advances in technology may reduce the need for standardization. Whereas standards for OCR in the past have dealt with such factors as fonts, materials, and character positioning, there is now an omnifont scanning device (the Kurzweil Data Entry Machine) that can successfully read typed or printed material with a mixture of fonts. Just as OCR devices are capable of processing a greater variety of printed input, devices for speech recognition are gradually becoming more versatile with advances in that technology.

Devices or media for information storage include punched cards, paper tape, magnetic disks, magnetic tape, microforms, and videodiscs. Compatibility of information storage devices is of particular concern when they are used as a basis for transferring information. Standards for magnetic tape are well established, specifying certain options for such characteristics as tape width and recording density. For information transfer to take place, suitable tape drives must be available at the receiving site. Microforms and videodiscs both have potential as large volume information storage devices. As the micrographics industry has matured, early tendencies toward proprietary design have given way to standardization. In contrast, videodiscs currently lack standards, with the result that there are a number of incompatible systems on the market. Development of videodisc standards is needed to insure interchangeability of data stored on discs. The result of an overabundance of incompatible systems is to hinder public acceptance of videodiscs as a storage medium.

Although standards for character sets exist, character sets developed for conventional data processing applications do not suffice for library use. It is necessary to develop standard ways of dealing with diacritics, non-Roman alphabets, and ideographic characters. Not only must standard ways of representing such characters be developed, but also input/output devices supporting such character sets must be available. Where a single standard character set does not exist, programs to convert from one representation to another must be written.

Facsimile, videotex, and teletex now supplement such familiar devices as the telephone and telegraph for information transmission. Although the newer technologies offer a variety of new capabilities and promise to support new types of information services, they also present problems due to the incompatibility of nonstandard equipment. As noted previously, there are two possibilities for achieving compatibility. In facsimile transmission, for example, one can choose identical machines as transmitters and receivers or develop mechanisms for conversion of signals, allowing communication between devices that would otherwise be incompatible. Large investments have been made by organizations in different countries in competing technologies for the encoding of graphics in videotex, making it difficult to reach agreement on an international standard.

A protocol is the set of rules governing the exchange of information between two systems. Protocols must be adhered to by both sender and receiver. They govern such things as mode, speed, character length and code which affect the format and timing of data transmission. Networks make use of standard protocols to establish connections between otherwise incompatible devices. Standardization of protocols in local area networks allows the interconnection of diverse types of office equipment. Unfortunately the local area networks developed by different manufacturers are frequently incompatible. For long haul data transmission, packet switched networks accommodate a wide range of often incompatible terminals and computers through standard protocols. Gateways and internetwork protocols allow global interconnection of packet switched networks. To interconnect networks of different types, such as videotex and packet switched, the gateway must perform conversion as well as switching functions. An important guide for co-ordinated standards development in networks is the ISO Reference Model for Open Systems Interconnection, which provides a common basis for co-ordination of standard protocol development for each of seven hierarchical layers from which a network is structured. Adherence to such a common set of protocols allows diverse types of computers to communicate. Although considerable progress has been made in establishing protocols for the lowest layers, the higher layers (presentation and application) still require additional work in standards development. While it may be possible to convert or map one protocol into another, the availability of standard protocols simplifies interconnection considerably.

6. NUMERIC DATA BASES

Throughout this report, the processing of bibliographic data has generally been assumed. However, information services have become increasingly involved in the processing of numeric data, and compatibility issues affecting the processing of numeric data bases (or what are frequently referred to as data banks, especially in Europe) fall within the scope of the present report.

Although numeric data can be considered more "concrete" than bibliographic data, they present much greater diversity, thus compounding the problem of compatibility. Sources of incompatibility can be found not only in the content of numeric data bases, but also in the systems for handling them, including data indexing techniques, interactive retrieval techniques, display formats, and computational and modeling capabilities (Fried/[166]). Negus/[359], for example, mentions one system that alone has 82 functions, each governed by a separate command, as well as 46 different statistical routines. Numeric data can also be represented in a computer system in a number of different ways. Luedke et al/[295], for example, distinguish among point representation, tabular representation, equation representation and graphic image representation. Numeric data are less stable than bibliographic data. There is little reason to change a bibliographic citation unless an obvious error is discovered in it. The values in certain numeric data bases, on the other hand, may change frequently. Moreover, the output from a numeric data base may be an output that is stored implicitly but not necessarily explicitly. For example, it can be an answer derived from some computation performed on the data stored. All of these factors complicate standardization in approaches to the handling of numeric data.

On the other hand, one could argue that consistency and compatibility are even more important in the processing of numeric data than in the handling of bibliographic data. As Kuipers and Thorpe/[274] have pointed out, the major reason for describing, representing, structuring and organizing numeric data is to permit their further manipulation. The data description and data structures adopted should be designed to facilitate further use of the data.

Because so many different ways exist to describe and structure numeric data, significant problems occur in the exchange or correlation of these data across different data bases. In general, in terms of compatibility and standardization, less has been achieved in the handling of numeric data than in the handling of bibliographic data. Luedke et al have stated:

> Incompatibility among numeric data bases and systems has up to now been a major obstacle to data exchange and intersystem communication. (Page 142)

Nevertheless, some progress has been made. Various recommendations concerning the presentation of numeric data (e.g., names, symbols and definitions used to represent quantity), especially their presentation in primary journals, have been made by the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions (ICSU), the International Union of Pure and Applied Chemistry (IUPAC), the International Union of Pure and Applied Physics (IUPAP), the International Union of Biochemistry (IUB) and the International Union of Nutritional Sciences (IUNS). The International Organization for Standardization has also produced several relevant standards (e.g., ISO-31, Quantities and Units), but, as Rigg/[407] has suggested, a complete standard would be a colossal undertaking.

A recent monograph sponsored by CODATA and Unesco presents an overview of data handling for science and technology (Rossmassler and Watson/418). An indexing problem of specific interest to the numeric data field is the augmentation of the primary scientific literature so that it is easier to find and retrieve numeric data (Eliezer/139). For this a set of numeric indexing principles is needed. Watson/527 has observed that:

> A given piece of numerical data generally refers to the magnitude of some quantity characterizing some property or phenomenon of a certain system, measured under a certain condition. The classification and indexing of data may be regarded as an elaboration of these facets. (Pages 134-135)

Approaches to data indexing include data flagging and tagging. A flag is a code symbol that indicates the presence of data. A tag is a more detailed identifier, normally used to characterize the measurement of something. Murdock/346 and Lerner/282 describe a number of experiments exploring the feasibility of data flagging and tagging in both the primary and secondary literature. In one recent study reported by Zaye/564, indexers for Chemical Abstracts Service assigned data tags to energy-related documents. These tags were alphanumeric codes of 2-4 characters, e.g., BP=boiling point, CD=current density.

To date experiments with flagging and tagging have focused primarily on the literature of a single discipline or the literature processed by a single abstracting service. The report of a Joint Working Group on Tagging and Flagging sponsored by CODATA and ICSU-AB/234 has cautioned that such independent development of schemes for flagging and tagging may lead to incompatibility. To accommodate searches for data in multidisciplinary areas, these schemes must be:

> (1) compatible with one another in the sense that no complete flag or tag has different meanings in different fields of knowledge;
>
> (2) unified in the sense that no concept is denoted by different combinations of characters in different fields of knowledge. (Page 19)

The ideal would be to have a scheme that is language independent, interdisciplinary, and internationally approved. This might necessitate changing some vocabulary and symbols well estabished in a particular discipline to terms that can be used across disciplines. The report describes possible tag codes ranging in length from 1 to 6 characters, the latter capable of denoting very specific kinds of numeric data and conditions under which measurements were made. Numeric data indexing has not yet become a common feature of bibliographic files because of the added costs of data tagging and the lack of a standard coding scheme familiar to indexers and searchers. Nolan and Whitney/373 propose an even more elaborate method for describing numeric data, presenting the syntax and structure of a language for precisely defining numeric data elements.

In numeric data bases per se, labeling of data is a possible source of inconsistency. Descriptors or short codes, which can be numeric, mnemonic, alphanumeric or expressed in natural language, act as markers to identify elements or clusters and to make explicit additional attributes of the data presented (Cronin/98). In physics, for example, such descriptors may include the units, data type, range of validity, and method of measurement (Behrens and Ebel/32). Another important attribute of data in the physical sciences is information on whether or not the data are critically evaluated, i.e., standard reference data (Berger and Tucker/38).

Some work on data element standardization has been performed. For example, Nolan and Whitney/[372] have discussed a "relational model" approach to standardization, with special reference to energy data, that would permit elements from different data bases to be combined within a composite data base. Such work can be considered as being at only a very preliminary stage. For certain data elements, of course, some well established standard can be followed. For example, it would usually make good sense to adopt Standard Industrial Codes (SICs) for the representation of industries within a data base.

An experiment underway on the DIALOG system should provide some insight into the usefulness of merged indexes as a means of accessing data values reported in various sources. SUPERINDEX/[248], a data base of back-of-the-book index entries of about 600 scientific, technical, and medical reference books, includes the indexes of CRC Press handbooks, the Merck Index, and Physicians Desk Reference, all of which include references to data. At present the handbooks themselves are not online; a record consists simply of an index entry giving page locations where the specified data appear. Although dealing with back-of-the-book indexes, results of this study should suggest problems likely to be found in merging schemes for data tagging and for data labeling as well.

In addition to describing data elements, it is possible to describe the data file as an entity. The need for such a description arises both in the context of creating bibliographic references to acknowledge use of a data file as an information source and in the context of cataloguing a data file so that its description can be included among the records representing a library's resources. Dodd /[122] has presented guidelines for creation of bibliographic references for numeric data files consistent with ANSI Z39.29-1977. In a later paper Dodd/[123] describes procedures for cataloguing machine-readable data files (MRDF) according to the chapter in AACR2 on MRDF and the MARC format for MRDF. Such data files would be separate entities on data carriers such as magnetic tape. Libraries of such files are found, for example, in social science data archives. The data elements specified for the MARC format for MRDF go beyond those required for describing monographs or serials. They reflect the special characteristics of this medium and the particular needs of file users. Dodd presents six levels of data elements:

(1) those needed to identify MRDF (e.g., bibliographic elements);

(2) those needed to describe the contents of MRDF (e.g., descriptive summary, data abstract, or in-depth subject analysis per item or variable);

(3) those needed to classify MRDF (e.g., appropriate classification codes, indexing or subject descriptors necessary to group like data files together);

(4) those needed to access MRDF (e.g., physical characteristics such as recording density and computer/software compatibility);

(5) those needed to analyze or use MRDF (e.g., citation of documentation and related reports, how/when the data were collected, unit of analysis, sampling procedures); and

(6) those needed to archive or maintain MRDF (e.g., in-house records pertaining to the processing, storage and use of the data file). (Pages 349-350)

Systems for handling numeric data online may be either generalized or tailored (Wanger and Landau/524). Generalized systems bring into one software package (or make links among several programs that are essentially transparent to users) capabilities to use one interface language in retrieving data, storing data, generating output, and performing statistical manipulations. These are command-structure interfaces. The user must learn the command required to perform desired functions and their formats. Tailored systems, on the other hand, provide particular capabilities or solve specific problems. They tend to be prompt-oriented interface languages, easier for the novice user but less flexible than the command interfaces of generalized systems. The tailored systems may be termed data base systems where data bases and software go together. Because of this coupling, different tailored systems are likely to appear highly dissimilar with little prospect for conversion between them. Even generalized systems may be quite diverse, with different data retrieval and manipulation capabilities.

Some work has been done in developing natural language interfaces which eliminate the need for the user to learn the command language required for access to a particular data base. Tennant/477 provides a number of case studies of implemented semantic analyzers providing natural language access to data bases, including the package INTELLECT which is available commercially. Marketed by Artificial Intelligence Corporation, INTELLECT has already been employed to provide access to a number of commercially available data base management systems widely used in industry. A separate lexicon must be built to characterize the content of each data base accessed. INTELLECT interprets the user's natural language queries and translates them into the formal command language of a given data base management system, reversing the process to answer the question in English. Hendrix et al/212 describe an even more ambitious prototype system named LADDER (Language Access to Distributed Data with Error Recovery) which provides natural language access to a large body of data distributed over a computer network. The user presents a query in a restricted subset of natural language and LADDER makes use of stored information on the names of specific fields, how they are formatted, how they are structured into files, and where they are physically located to provide an answer. These programs for natural language access currently work only in limited subject domains but, where successful, they can shield the user from incompatibilities in file structure and access methods.

Many of the numeric data files available online deal with chemical information. In chemistry, the Chemical Abstracts Service registry number has proved valuable as a unique identifier for chemical substances, allowing unambiguous linking between files. Callahan and Rusch/71 describe the use of the registry number in linking the Chemical Abstracts Service registry nomenclature file to the CA SEARCH bibliographic file on DIALOG. Heller and Milne/206, 207 describe the use of registry numbers in the NIH-EPA Chemical Information System (CIS), a joint project of the U.S. National Institutes of Health and the Environmental Protection Agency. Using the registry number, one is able to go back and forth between CIS components to obtain various properties of a compound, such as physical and toxicological data. Users with intelligent terminals can take the registry numbers from a CIS search and store them for input to other online systems such as DIALOG. Because of the modular nature of CIS, more components can be added if the registry numbers are available for linking. One of the attributes of a chemical substance which is frequently of interest is its structure. Jacobus et al/251 have described the problems in achieving compatibility of representations for chemical structure.

Large organizations frequently maintain many data files, some of which may overlap in content. Two U.S. government agencies have recently explored the value of a data element dictionary/directory to manage data. Such a directory, identifying and describing the data elements found in each file, serves not only as a repository of information about an agency's data, but also as a vehicle for standardization of data elements where overlap is found. Travis/[487] describes work on the Energy Information Agency Data Resources Directory and Melley/[327] explains a similar project within the Environmental Protection Agency.

Data have proliferated with the rapid extension in power and versatility of data creating and recording equipment. Because one computer may be used for data capture and another for data analysis, data portability becomes a concern. Bouillé/[47] has observed that data portability requires use of a data structure model that is completely independent of the concept of storage structure. He describes one such data structure model. A draft standard for the exchange of numeric data has been formulated within the United States by the Interlaboratory Working Group for Data Exchange (IWGDE), within the Department of Energy. The draft defines a generalized structure that can be used by systems to exchange records representing a wide and open-ended variety of data types and structures. The draft standard, a generalization of the ANSI Z39.2-1971 and ISO 2709 bibliographic interchange formats is already being implemented by various laboratories in the United States.

Merrill and Austin/[329] have envisaged the development of "resource data base systems" that will directly access a wide variety of data banks in many locations and allow their interrogation through a common command language. Nevertheless, Negus/[359], in his feasibility study, concluded that a common command language for numeric data systems appears unpromising: because numeric data systems support not only retrieval but also manipulation of data and presentation of the output in a variety of formats (e.g., charts, tables, graphs), there are many sources of incompatibility among these systems.

Because of the growing availability of numeric data bases, there is a need for increased attention to factors affecting compatibility of both data representations and the systems for handling data. Some progress has been made in schemes for handling data tagging and flagging and in codes for cataloguing machine readable data files. However, systems for handling numeric data exhibit considerable diversity in approaches to data retrieval, manipulation, and display.

7. MANAGEMENT ASPECTS

Compatibility is most often thought of as applying to technical activities, such as cataloguing, and to various applications of technology. Although rather more nebulous, compatibility in some managerial and procedural aspects of information services can also be recognized.

To begin with, professional education imposes a type of compatibility within professional practice, at least at national levels. In the United States, for example, many schools of library and information science exist. The substance of what is taught, however, is much the same from one school to another. The graduates of different schools can work effectively together because they share a common corpus of knowledge, terminology and even, perhaps, attitudes. As a result, their approaches to the implementation, organization and management of information services are likely to be compatible.

Large information systems go further through continuing education. Thus, international systems such as INIS and AGRIS, besides establishing standards, promote compatible operations and even operating philosophies through periodic workshops, newsletters and other devices.

Procedural compatibility is achieved in additional ways. Indexing manuals, for example, tend to ensure that different indexers, possibly widely distributed over many different centres, use similar approaches in analyzing items. A document such as Unesco's *Indexing Principles* (SC75/WS/58, September 1975) thus serves to impose some general level of compatibility in indexing philosophy and procedure, independent of any particular vocabulary or "system".

The managers of a network can seek some level of administrative compatibility through, for example, the use of standard forms and the development of procedural guidelines. The Commission of the European Communities, responsible for the operation of Euronet, presents a notable example: guidelines for co-operation between data base suppliers and computer hosts have been prepared, a model agreement between a host and a host customer has been drawn up, and work has begun on the harmonization of billing and revenue collection procedures. The harmonization of user manuals and of approaches to the education and training of users have also been investigated for the Euronet-DIANE situation (Casey/72).

Another notable example is the adoption of standard forms and procedures in the exchange of photocopies or the lending of books among libraries. All such standardization activities can promote efficiency and reduce costs. They also tend to ensure the transferability of personnel, at least within national boundaries. That librarians and other information specialists can usually move easily from one institution to another is at least partly attributable to the existence and adoption of common procedures and protocols.

A number of standards relating to various types of library service exist and several of these are listed in the *UNISIST Guide to Standards for Information Handling*. Such standards are tailored to particular types of libraries--public, school, university, hospital, and so on. They tend to be quantitative standards--e.g., how many books, how much seating--rather than qualitative. Consequently, they relate more to inputs to libraries than to their outputs (services), which are much less easily quantified. Over the years, this has been one of the major criticisms leveled at standards of this type. These standards have little transferability across national boundaries. In fact, they are usually developed by looking at the characteristics of libraries that are assumed to be good examples of their type. It follows, then, that a standard for college libraries in the United States has little real relevance to a college library in, say, an African nation, where size and composition of the student population, as well as the curriculum, are likely to be quite different.

Another problem with library standards is that they are usually intended to be *minimum* standards. This is not always recognized, particularly by those responsible for funding. Regrettably, some fine libraries have found such standards to work against them. The fact that these institutions far exceed the standards, creates a justification (misguided) for reducing their funding levels.

In general, standards governing various facets of information processing do not have the force of law so it is not possible to require their use.* If they are observed, it is because the organizations involved choose to observe them. The most that can be imposed in the way of enforcement, in most cases at least, is to make the adoption of certain standards a condition for participation in some co-operative endeavour, such as a network.

In the case of administrative concerns, however, actions may indeed be governed by laws, regulations or international conventions. Obvious examples occur in such areas as copyright and transborder data flow. In such cases, observance of legal instruments or treaties ensures standardized behaviour and, in effect, imposes a type of compatibility on the way institutions or even nations process certain types of information. Nevertheless, incompatibility among various national laws does hinder effective exchange of information. The Organization for Economic Cooperation and Development (OECD) and the Council of Europe have undertaken to harmonize these laws. OECD approved voluntary guidelines for the protection of privacy in 1980.

---

* This situation is quite different from that of manufacturing, where a customer is able to demand that a manufacturer adhere to a particular set of standards.

8. CASE STUDIES

8.1 Introduction

The purpose of these case studies is primarily to give concrete examples of how various types of organizations have achieved some level of compatibility to facilitate the exchange of information. An attempt has been made to vary the examples, in order to include relatively simple approaches as well as highly sophisticated ones, and to illustrate a number of the facets of compatibility--records formats, bibliographic description, index languages, software, and so on--discussed earlier in the report. At the end of each study references are made to the earlier sections to which the study relates.

8.2 The Regional Bibliographic Centre, Standing Conference of Pacific Libraries (SCOPAL)*

An example of compatibility issues affecting co-operative bibliographic processing on a modest scale can be found in the activities of the Regional Bibliographic Centre of the Standing Conference of Pacific Libraries. The centre was established in 1979 at the University of the South Pacific in Suva, Fiji. The major objectives of this centre include: producing, in printed and machine-readable form, regional bibliographies (works about the region as well as those produced within the region) for monographs and periodical articles; assisting individual countries in producing their own national bibliographies; and promoting appropriate standards for the collection of the bibliographic data.

A simple input form (allowing cataloguing to the first level of AACR2) is now being used by libraries in the region to supply bibliographic data to the Regional Bibliographic Centre. Both regional and national bibliographies are produced through co-operative effort; participating countries derive national bibliographies from the joint project when, in fact, some of them had not previously been able to produce such tools. Some of the national bibliographies are issued separately by individual countries (e.g., Papua New Guinea produces for Nauru and Norfolk Island, Hawaii produces for American Samoa and the American territories in Micronesia). The regional bibliography is computerized and output can be generated for the printing of national as well as regional lists.

That such co-operation can occur is very largely due to the existence of appropriate standards and the willingness of the various institutions to adopt them. International standards endorsed by SCOPAL in 1978 and 1979 include: AACR2, the ISBDs, Library of Congress subject headings, and the arrangement of entries in national bibliographies by the Dewey Decimal Classification. The MARC format is also followed in the region. The acceptance of international standards, besides facilitating co-operation (e.g., in various types of compilation and processing activities), also tends to ensure that the various printed bibliographies and union catalogues produced in the region are compatible in terms of arrangement and general format.

This case study illustrates the following sections of the report: 2.1, 2.2

8.3 DEVSIS

DEVSIS is the name given to a proposed international information system for the development sciences. In 1975 a DEVSIS Study Team, sponsored by IDRC, ILO,

---

* In 1982 the Regional Bibliographic Centre was renamed the Pacific Information Centre.

OECD, ESA, UNDP and Unesco, prepared a preliminary design for DEVSIS/[113]. Pointing out that "development is one of the most important missions facing the world" yet "no global systematic attempt has yet been made to provide information support to those responsible for economic and social development", the study team recommended a worldwide DEVSIS system modeled after two information systems, INIS and AGRIS, already existing within the UN family. As in the INIS/AGRIS model, each country and each international body would accept responsibility for identifying, collecting, cataloguing and indexing development-related literature produced within its own territory or organizations. A central unit within the UN would merge these inputs and produce outputs in the form of searchable magnetic tapes and printed indexes.

In point of fact, DEVSIS has never been implemented as such a global system. Nevertheless, some elements of a global DEVSIS do exist. For example, IDRC is now building a data base of development literature produced in Canada, the Federal Republic of Germany, and elsewhere. A printed *Devindex*, providing a guide to the Canadian development literature, was produced for the 1975 and 1976 material; the 1977 issue was extended to include the German material, and the 1978 issue includes development literature from Indonesia, Morocco, the Netherlands and the Soviet Union, as well as Canada and the Federal Republic of Germany. The scope continues to be expanded by the addition of further countries.

The IDRC data base is regarded as an ongoing experiment. The existence of the data base allows IDRC to examine, test and evaluate methodologies for information exchange and resource sharing. Several modifications of the DEVSIS methodology have been proposed, as a result of experience with this data base, and have been accepted by various other organizations involved in the processing of development literature.

The first issue of *Devindex Pakistan*, covering the 1976 Pakistani literature on economic and social development, was published in 1978 by the Pakistan Institute of Development Economics. Similar work is being performed in the Philippines, Thailand, Bangladesh, India and Australia. IDRC has assisted this work in various ways (e.g., by processing of worksheets).

Various international agencies, especially the regional economic commissions of the United Nations, can be considered as essential components in what is becoming a loosely structured DEVSIS network. In Latin America, CLADES is building a data base of ECLA (i.e., CEPAL) and ILPES documents going back to 1970/71. CLADES was established as a unit of ECLA in 1971. The functions of CLADES include technical assistance in documentation, training, co-ordination, research and studies, and dissemination of information through publications and the organization of meetings. With the financial assistance of IDRC, CLADES and ILPES have worked since 1979 on a joint project to create INFOPLAN, an information system in the area of planning. The publication PLANINDEX is a product of INFOPLAN.

ECLA is also active in the Caribbean. The Caribbean Documentation and Information Centre (CDIC) directs a programme designed to link existing information units into networks, the first being the Caribbean Information System for Economic and Social Planning (CARISPLAN). Other networks are planned for agriculture and for science and technology. CARISPLAN is a decentralized system; the focal points are the information units of the various national planning agencies (fourteen countries now participate). The CDIC provides a number of services, including training, prints the *CARISPLAN Abstracts* and works to strengthen the national information units and to ensure compatibility among them. Other services provided include bibliographic searching, document delivery, and the production of a current awareness bulletin. At present, all computer support is provided from ECLA in Santiago. INFOPLAN can be considered a regional system built on the DEVSIS model, while CARISPLAN can be considered a sub-regional component of the DEVSIS network. Detailed recommendations regarding a common bibliographic format for use within the Caribbean region are given in a report by Dierickx/[117].

In Europe, the OECD Development Centre maintains various data bases. Its parent, OECD, is the publisher of the *Macrothesaurus for Information Processing in the Field of Economic and Social Development* and maintains a machine-readable version of this tool.

In Africa, ECA maintains the PADIS project, which has or will have a number of integrated components: PADIS/DEV (a data base of development literature from Africa); PADIS/COM (complementary sectoral data bases on such subjects as human settlements, mineral resources, population, energy, and women); PADIS/STAT (statistical data relating to Africa); and PADIS/NET, a telecommunications network designed to make the data bases of the developed countries accessible to the African nations. A major product of PADIS is the printed *DEVINDEX/Africa*.

ECWA has plans to establish an Arab regional documentation centre for economic and social development. Both bibliographic and statistical information are to be included.

The situation at ESCAP is more diffuse: about a dozen separate information systems - in population, trade, trade statistics, and other subject areas, as well as the ESCAP Library - exist, with little co-ordination among the various units.

IESA maintains in New York an Information Systems Unit (ISU) which collects and indexes unpublished reports arising from UN activities in economic and social development (including the subject of human settlements). A publication, *Development Information Abstracts*, is produced and retrospective searches are performed on demand.

These achievements have profound significance for the subject of this report. The original concept of a global DEVSIS has disappeared and been replaced by the adoption, among the various DEVSIS components, of common methodologies to allow free interchange of information. In other words, a DEVSIS network has emerged in the form of several distinct centres all using common or compatible information processing methodologies.

Compatibility has been achieved at a number of levels:

1. Adoption of the *Macrothesaurus* as a standard for subject indexing. In some cases, the *Macrothesaurus* has been expanded to more fully meet the needs of a particular centre. For example, the Information Systems Unit of IESA has found it necessary to add a substantial number of terms relating to human settlements, and IDRC has also added new terms. Input from the various centres, in theory, serves as an important means of keeping the thesaurus up to date. Moreover, various other thesauri produced within the UN have been made as compatible as possible with the *Macrothesaurus*. One example is the *Environment Thesaurus for Latin America and the Caribbean*, published by CLADES. The *UNBIS Thesaurus*, developed for use by the Dag Hammarskjold Library, also claims to be compatible with the *Macrothesaurus*, although, clearly, it is much broader in scope. The *Macrothesaurus* is officially trilingual (English, French, Spanish) and an Arabic version is now being worked on. Thus, indexing can be in any one of these languages and retrieval can use any language.

2. In general, the UNISIST *Reference Manual* has been adopted as a standard for the structure and content of bibliographic records. Again, however, some centres have found it necessary to introduce further data elements.

3. The ISO-2709 format is followed for transmission purposes.

4. Compatible worksheets for cataloguing and indexing have been developed. The worksheet has been progressively refined over the years. The most recent version, that developed for PADIS, seems to be becoming a kind of network standard.

5. Processing software adopted by IDRC, the ISU and ECA/PADIS is MINISIS, a minicomputer version of the ISIS software originally developed by ILO. MINISIS was compiled by IDRC and this software is made freely available by IDRC to agencies of the UN and to developing countries. It is well suited to the developing countries because it is designed for use on relatively inexpensive minicomputers. MINISIS will accept input of bibliographic records, "manage" the file of these records, permit the file to be searched, and generate various forms of system output including publication formats. Other components of the DEVSIS network use the compatible ISIS software.

6. Some measure of hardware compatibility also exists. Hewlett-Packard minicomputers (specifically the series 3000) are used by IDRC, the ISU and ECA.

At a Technical Expert Group Meeting on Common Indexing Tools, held in New York in June 1981, the value of exchanging or merging the various component data bases was recognized. It was recommended, therefore, that IESA/ISU make its data base available to each of the regional commissions, that the regional commissions make their data bases available to IESA/ISU, and that IESA/ISU create a merged data base, possibly to be re-distributed to each of the commissions. In 1981, the ISU made copies of its DIS data base available to ECA, CEPAL, ECWA and ICC (Geneva). In exchange, a copy of the CEPAL data base was received by the ISU.

That such interchange has been possible is due entirely to the compatibility deliberately sought and achieved by the various components of this DEVSIS network.

Despite these achievements, the DEVSIS situation is not free from problems, especially in the use of the *Macrothesaurus*. While OECD maintains the tool for its own use, it has not assumed the responsibility for network-wide maintenance and updating. Consequently, a number of variants, including versions in different languages, now exist. The New York meeting recommended that IESA/ISU, in co-operation with the regional commissions and other users, assume the responsibility for continuous updating of the thesaurus. This would include using its minicomputer facilities for processing modifications and the distribution of these modifications to all users. The ISU has accepted this responsibility and a new edition of the thesaurus is planned for 1983.

Work to promote compatibility within DEVSIS continues. A two-week meeting was convened by IDRC in Canada in November 1981. The meeting was attended by representatives of the regional economic commissions and other regional institutions interested in exploring the DEVSIS model. An important outcome of this meeting is the publication by IDRC of *Recommended Methods for Development Information Systems*. The first volume, June 1982, deals with bibliographic description and record preparation/[338]. Two further volumes will deal with procedures for abstracting and for input to computerized authority files.

This case study illustrates the following sections of the report: 1, 2.2, 2.3, 3.1, 4.3, 5.1.

8.4 TOXLINE and POPLINE: a case study in the merging of data bases

TOXLINE and POPLINE are two online data bases maintained by the National Library of Medicine (NLM) in the United States. They are treated together because they offer some interesting contrasts in the solution of problems associated with the merging of data bases. These problems have been identified by the National Library of Medicine as follows:

1. Differences in record formats.
2. Identification and elimination of duplicate records.
3. Different types of indexing.
4. Proprietary or copyright protection of the records.

All but the last of these involve compatibility issues.

TOXLINE (Toxicology Information Online) was established in 1972. Before that time the literature of toxicology was scattered over many different data bases. TOXLINE was conceived of as a unified data base drawing subsets of toxicology from other secondary sources. Bibliographic records are acquired from these sources in various ways:

1. Through purchase from a data base producer
2. Through leasing (requiring payment of royalties)
3. Through inter-institutional agreements (which may also involve some form of payment)
4. From other National Library of Medicine data bases.

All records, from whatever source, are converted to a standard TOXLINE format.

In 1981, TOXLINE was inputting records from nine sources:

1. Chemical Abstracts Service: Chemical-Biological Activities (CBAC)
2. Health Effects of Environmental Pollutants (HEEP) from the Biosciences Information Service (BIOSIS)
3. International Pharmaceutical Abstracts of the American Society of Hospital Pharmacists (ASHP)
4. NLM's own Toxicity Bibliography (TOXBIB), derived from MEDLINE
5. Pesticides Abstracts (PESTAB)
6. The Environmental Mutagen Information Center (EMIC)
7. The Environmental Teratology Information Center (ETIC)

8. Toxicology/Epidemiology Research Projects (RPROJ) of the Smithsonian Science Information Exchange (SSIE)

9. The Toxicology Document and Data Depository (TD3) of the National Technical Information Service

SSIE is no longer a current source of records. However, it is hoped that the National Technical Information Service (NTIS) will be able to provide updates to RPROJ in the near future.

The TOXLINE data base now contains over 1.2 million bibliographic references to published human and animal toxicity studies, effects of environmental chemicals and pollutants, adverse drug reactions, and analytical methodology. The data base, which contains abstracts for most items, is growing at about 140,000 records a year.

Table 4 shows the 23 data elements (plus textwords) appearing in a TOXLINE record and indicates whether or not each element is present in the various component data bases. The country or state designator (CY), for example, appears only in RPROJ and CBAC records, while the language designator (LA) appears in all but two of the source files. The table also shows which data elements are directly searchable online ("search status") and which are printed out in response to various online print commands: "print", "print full" and "print detailed".

One problem encountered by the manager of a merged data base is the fact that the component data bases are not necessarily completely stable as to data elements and record formats. If a component data base is altered - e.g., a data element added to or deleted from the record - the program that converts that record format to the format of the merged data base will need to be modified. Clearly, the more component data bases, the greater this problem is likely to be.

With eight or nine sources providing entries to TOXLINE, a substantial possibility exists that records for the same bibliographic item will be derived from two or more of the sources. The handling of duplicates is a significant problem in the management of a merged data base drawing on a number of sources. In TOXLINE, search keys have been developed to identify duplicate records. The component data bases are ranked to reflect a set of priorities for the retention of records. When duplicates are encountered, the ranking determines which record is retained in the TOXLINE data base. In general, when the information content is equivalent, records from sources not requiring use payments are given preference over those from sources that do.

A major problem in the handling of duplicates is the fact that, since different data bases have different priorities and throughput times, the duplicate records do not arrive at the same time. Consequently, it is not efficient to try to eliminate duplicates on a continuous basis. Instead, the data base is purged of duplicate records by periodic "rejuvenations"; duplicate records will appear in the online files in the period between such rejuvenations.

The various component data bases of TOXLINE follow different indexing and vocabulary conventions. Some use controlled vocabularies while others rely on natural language. No attempt is made to impose any additional vocabulary requirements on these data base producers. Moreover, no attempt is made to apply any type of normalization to the various vocabularies. Instead, all vocabularies (controlled or not) are treated as uncontrolled terms and put in the Keyword field. In addition, all of the component words (i.e., Uniterms) appearing in this field, along

TOXLINE MATRIX

August 14, 1981

category qualifier	element name	search status	print	print full	print detailed	* * * * * * * * secondary files * * * * * * * * * * *								
						toxbib	cbac	pestab	ashp	heep	emic	etic	rproj	td3
ab	abstract	tw		ab	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab
ad	address			ad	ad	ad	ad	ad	ad	ad			ad	ad
au	author	*	au	au	au	au	au	au	au	au	au	au	au	au
aw	award type	*			aw								aw	
cd	coden	*			cd	cd	cd	cd	cd	cd	cd	cd		
cy	country or state	*			cy								cy	
dtg	data tags	*			dtg		dtg							
em	entry month	*			em	em	em	em	em	em	em	em	em	em
id	identification number	*			id								id	id
is	international standard serial number	*			is	is								
kw	keywords	*tw			kw	kw			kw		kw	kw		kw
la	language	*		la	la	la	la	la	la	la	la	la		
od	order number			od	od									od
pr	price			pr	pr									pr
rn	cas registry number	*			rn	rn	rn	rn	rn	rn	rn	rn		
sa	supporting agency				sa								sa	sa
si	secondary source id	*	si	si	si	si	si	si	si	si	si	si	si	si
so	source		so	so	so	so	so	so	so	so	so	so	so	so
ti	title	tw	ti	ti	ti	ti	ti	ti	ti	ti	ti	ti	ti	ti
tw	text words(from ab,kw,ti)	tw				tw	tw	tw	tw	tw	tw	tw	tw	tw
y1	initial year				y1								y1	
y2	final year				y2								y2	
yr	year (part of source field)	*				yr	yr	yr	yr	yr	yr	yr	yr	yr
zp	zip code	*			zp								zp	

&cgm.tox.matrix

Table 4.

Data elements in the component data bases of TOXLINE.

with the component words of the title and abstract field, are extracted and put in the Text Word field, where they are searchable. In other words, vocabulary "compatibility" is achieved by treating everything as natural language.

Some further problems encountered in the development of TOXLINE have been discussed by Schultheisz/435. While data have generally been supplied by the data base producers in magnetic tape form, the tapes themselves have varied from unlabeled 800 BPI to labeled 1600 BPI. Record formats and blocking factors have varied from unblocked to variable blocked spanned records. In some cases, records had to be input by NLM from a printed publication because no machine-readable records were provided by the publisher. In the data supplied, numerous structures and formats are evident. Some are unstructured in the sense that data elements are not explicitly defined (making them difficult to convert to a single standardized format) while others are in the form of structured records with highly organized data in specified data elements. Another problem is that a data base created to generate a publication is likely to contain various machine codes that are not useful in the online file and need to be stripped away.

TOXLINE, then, presents an excellent example of the problems that must be faced in creating a useful tertiary data base derived from a number of component sources. As Schultheisz points out:

> Although any single program may have been trivial to design and write, when all conversion program requirements were identified and the entire task was reviewed, the work was quite formidable. (Page 424)

A second merged data base maintained by NLM is POPLINE (Population Information Online) which covers the literature of family planning, fertility control, population and reproduction. POPLINE was formed from, and continues to be supplied by, two major data bases: POPINFORM, a data base of the Population Information Program (PIP) at Johns Hopkins University, and the data base of the Center for Population and Family Health (CPFH), Columbia University. More recently, a third component has been added: the Population Index of the Office of Population Research, Princeton University. The POPLINE unit record is more complex than that of TOXLINE. It contains 40 data elements and is illustrated in Table 5.

Although PIP and CPFH have redesigned their input (indexing) forms to make them more compatible with NLM requirements, conversion programs are still needed to convert input from these centres to the required NLM processing format (ELHILL). In point of fact, input from both centres goes to a contractor who puts it into "card image format". When the contractor tape reaches NLM it must go through a conversion program to put the records into ELHILL format. Input from Princeton also goes through a conversion program but is handled as a separate operation.

As of 1982, the POPLINE data base includes approximately 100,000 bibliographic records and is growing at the rate of about 14,000 each year.

POPLINE presents an interesting contrast with TOXLINE. In the first place, it is a less complex operation because it involves many fewer component data bases. Moreover, the POPLINE participants have divided up the population literature by mutual agreement, so the problem of duplicate records is virtually eliminated.

It is in the handling of the indexing vocabulary, however, that POPLINE differs most from TOXLINE. All three centres now index using a common thesaurus,

Table 5

The POPLINE unit record

Category Qualifier	Element Name	Search Status	PRINT	PRINT FULL	PRINT DETAILED	PRINT AR
AB	ABSTRACT	TW			X	X
AD	ADDRESS			X	X	
AU	AUTHOR	*	X	X	X	X
CN	CORPORATE NAME	*, TW	X	X	X	X
CO	COLLATION				X	
CY	COUNTRY	*			X	
DA	DATE OF ENTRY	*, R			X	
DP	DATE OF PUBLICATION	*			X	
EM	ENTRY MONTH	*			X	
GN	GENERAL NOTES	TW		X	X	
IM	IMPRINT				X	
IP	ISSUE/PART/SUPPL				X	
IS	ISSN	*			X	
JC	JOURNAL TITLE CODE	*			X	
KW	KEYWORDS	*			X	
LA	LANGUAGE	*		X	X	
LR	LAST REVISION DATE	*, R			X	
MH	MeSH HEADINGS	*		X	X	
MI	MONOGRAPHIC INFORMATION	TW			X	
MN	MeSH TREE NUMBER	*				
MR	MAJOR REVISION DATE	*, R			X-	
NM	NAME OF SUBSTANCE	*		X	X	
PG	PAGINATION				X	
PS	PERSONAL NAME AS SUBJECT	*			X	
PT	PUBLICATION TYPE	*			X	
RF	NUMBER OF REFERENCES				X	
RI	REVISION INDICATOR				X	
RN	CAS REGISTRY NUMBER	*		X	X	
RO	RECORD ORIGINATOR	*			X	
SE	SERIES TITLE	*, TW			X	
SI	SECONDARY SOURCE ID	*	X	X	X	X
SO	SOURCE		X	X	X	X
TA	TITLE ABBREVIATION	*			X	
TI	TITLE	TW	X	X	X	X
TJ	TITLE OF JOURNAL	TW			X	
TT	TRANSLITERATED/VERNACULAR TITLE				X	
UI	UNIQUE IDENTIFIER	*			X	
VI	VOLUME ISSUE				X	
YR	YEAR	*				
ZN	MeSH Z TREE NUMBER	*			X	

* = Directly Searchable
R = Rangeable
TW = Text Word Searchable

[1] The Source (SO) field can be composed of the following data elements dependent on publication type: Title of Journal (TJ), Monographic Information (MI), Date of Publication (DP), Imprint (IM), Volume Issue (VI), Issue/Part/Suppl. (IP), Pagination (PG), Collation (CO), and Series Title (SE).

the POPLINE Thesaurus. This was formed by the integration of two separate thesauri previously used by CPFH and PIP, with additional terms drawn from other population sources, including the Population Multilingual Thesaurus and NLM's own Medical Subject Headings (MeSH). Further demographic terms were added to take care of the needs of the Population Index.

Both CPFH and PIP enter records for a substantial number of items that have also been indexed into NLM's major data base, MEDLINE. Unfortunately, MEDLINE records cannot be transferred intact. While some MeSH terms are exactly equivalent to descriptors in the POPLINE Thesaurus, the population-related vocabulary of this latter tool is much more specific than the population-related vocabulary of MeSH, necessitating double indexing for POPLINE and MEDLINE purposes.

This case study illustrates the following sections of the report: 2.2, 3.1, 3.3, 5.3.

8.5 The exchange of bibliographic data: a case study in geology

The international cooperative systems, such as INIS and AGRIS, exemplify the construction of a single machine-readable data base having input from a number of different centres. Compatibility in this case is achieved by various requirements and standards observed by all the participating centres.

A different approach to constructing a composite data base is to draw bibliographic records, in some defined subject field, from a number of pre-existing data bases, and to merge these in a tertiary data base. Some measure of compatibility is imposed on the records by the producer of the tertiary data base.

A third possibility for the merging of bibliographic records also exists. It involves two or more independent data base producers agreeing to exchange records in some subject of common interest. One example of this type of co-operation exists in the field of geology as the result of an agreement signed in 1980 by the American Geological Institute (AGI), the Centre National de la Recherche Scientifique (CNRS) and the Bureau de Recherches Géologiques et Minières (BRGM). These organizations, respectively, produce three data bases: GeoRef, PASCAL (which includes geology) and GEODE. The main objective of the agreement is to improve worldwide coverage and control of the literature of geology.

The existing data bases continue in the base languages, English and French, but each now contains references in the composite data base. Two printed bibliographies, the Bibliography and Index of Geology and the Bulletin Signalétique--Bibliographie des Sciences de la Terre, are generated from the composite data base. Coverage includes all materials previously indexed into the component data bases, with AGI contributing the North American literature and BRGM-CNRS the European, literature from other areas being divided between the institutions. Duplication of processing is thus eliminated, while all three data bases remain worldwide in coverage and each includes all references involved in the exchange.

Records are exchanged in machine-readable form. Before this exchange could begin, differences in indexing, data elements and tape formats had to be reconciled. This required some compromises on both sides.

Writing from the viewpoint of the American Geological Institute, Mulvihill/345 discussed some of the problems that had to be faced in putting the French records into the AGI format:

...the AGI computer has to be able to put the data from BRGM-CNRS into the AGI format. This meant not only that each data element used in PASCAL-GEODE should fit into a data element in GeoRef. It also meant that the computer should be able to use the actual data as provided or be able to convert that data to the required form. To give one example, if PASCAL-GEODE recorded the language of the source document, then the AGI computer should be able to put that information somewhere in GeoRef. Moreover, there must be agreement on whether the language will be a symbol or spelled out, be in upper or lower case letters, be repeatable or not, etc. If the language comes as a symbol and AGI wishes to convert it to another symbol or to full form, then the computer needs a complete list of the symbols used. ...Such are the details the computer forces upon one. This is good but requires some considerable work on specifications and programming. To minimize this effort we decided to establish an exchange format. AGI might have any sort of internal format it wished but, for the exchange, the data must be put into the exchange format. Similarly, any data coming to AGI had to be put into the exchange format. For the exchange format we decided upon an existing standard, the UNISIST-ICSU/AB *Reference Manual for Machine-Readable Bibliographic Descriptions*. The Reference Manual was intended to be used as an exchange format. Adoption of this exchange format gave our programmer detailed, ready-made specifications to which and from which to convert data. Even so, each data element had to be discussed and specifics filled in tailored to our needs. This application of the Reference Manual format was written out and dubbed the Atlas format. Extensive programming was required, but the common exchange format at least gave the programmers detailed specifications to work from.

...Ideally AGI should not have to look at the references received from BRGM-CNRS before adding them to GeoRef. For this to be done there would have to be automatic checking in of references in serial and monograph files in the library, no possibility of duplication of references, confidence in the completeness and accuracy of the data, machine translation of index terms, titles and abstracts, and incorporation of all data into the GeoRef format by computer only, without human intervention.

Unassisted addition of records has not been achieved and probably never will be. However this principle guided us toward using the computer for all feasible tasks and attempting to eliminate differences between us. For example, if we could decide on a common set of category codes and apply them consistently, then the AGI indexer reviewing a reference from BRGM-CNRS would not need to assign a category code to the reference. Past practices and the decision by both parties not to change their printed bibliographies worked against a common set of categories, but a compromise was achieved.

...By far, most of the weeks spent working out problems were taken up with indexing terminology. This was so important because if an exchanged document had to be completely reindexed, much of the savings of the exchange would be lost. Further, since AGI might not have the source document from which BRGM-CNRS had prepared the references, at the minimum there should be enough indexing provided in that reference for the AGI indexer to work from. So the indexing of areas,

> commodities, elements, rocks, minerals, fossils, etc. was discussed. For each of these key topics, indexing practice and terminology were compared and what each party expected from the other was established. A second phase involved a term-by-term comparison of the two vocabularies, complicated by the involvement of two languages. The questions what do you mean by such-and-such a term and how do you use this term were asked over and over. We tried to keep changes to a minimum, and where a change was made, to make it in a way that would be compatible with our previous indexing. (Pages 3-4)

Some other changes in GeoRef result from the co-operation, including:

1. The base language is extended to English plus French. Titles, index terms and abstracts (if any) appear in both languages.

2. A bilingual table of equivalencies, in machine-readable form, translates index terms from French to English (and vice versa) automatically.

3. Diacritical marks (not previously used) had to be introduced for some European languages.

4. Some new data elements had to be introduced into the GeoRef record.

5. A translation step had to be introduced. Titles, abstracts and uncontrolled index terms are translated by AGI from French into English and these translations are also sent to France. The equivalent translation step occurs in France for English material supplied by AGI.

6. The GeoRef Thesaurus was expanded to include new terms necessitated by the exchange.

7. Conventions had to be agreed upon to take care of such matters as capitalization, place names and translations.

Clearly, equivalent changes had also to be made by the co-operating organizations in France. Michel/[332] has pointed out that CNRS was involved in extensive reformatting and reprogramming to make the exchange possible from the French side. Further information is given by Rassam et al/[403].

This case study illustrates the following sections of the report: 2.2, 2.3, 3.1, 3.7.

8.6 CCL: a case study in standardization of command languages

The Common Command Language (CCL) is a high level command language in which each operation comprises a command word followed by one or more parameters entered according to simple rules. Retrieval systems using the CCL, which means following the recommended syntax and semantics and not just using CCL command names, present the user with an easily learned user interface. However, standardization of commands does not mean that all systems are identical. Response to a command, both in terms of information presented to the user and in terms of actual operations carried out, may differ from system to system or even from data base to data base within one system.

Exploration of the feasibility of developing a common command language occurred as Euronet-DIANE (Direct Information Access Network for Europe) was being planned. Euronet has a relatively large number of hosts with outwardly different retrieval systems each providing access to a comparatively small number of data bases. Thus many users are likely to find a need to access two or more systems regularly and perhaps several others on occasion (Negus/357). Recognizing that diversity in command languages was likely to inhibit full use of the resources to be made available online through Euronet, the Commission of the European Communities employed Alan Negus to carry out a feasibility study during 1976.

Feasibility was evaluated at both the conceptual and technical levels. At the conceptual level, a study of 11 systems led Negus to conclude that the functions provided by different retrieval systems were more or less similar. Furthermore, this set of common functions is sufficiently flexible to meet user needs and to provide access to a variety of bibliographic data bases. At the technical level, two possibilities were identified for implementation of CCL by Euronet hosts. Either existing retrieval systems could be modified to accommodate the CCL or a preprocessor could be used to interface CCL commands with those already used by the host. Given this feasibility assessment, Negus prepared the recommendations for a CCL which were issued in December 1977. This document, titled EURONET Guideline: Standard Commands for Retrieval Systems/358, presents a formal definition of CCL which has been used as the basis for its implementation by a number of Euronet hosts. For each command defined, the formal definition of syntax is followed by some typical examples and narrative description (semantics) of the effect and intended use of the command. The CCL aids easy transfer from one host to another, thus implementing the common-information-market policy basic to Euronet-DIANE.

Specification of a common command language is feasible because most bibliographic retrieval systems supporting Boolean search strategies and inverted files share a number of common functions, even though the actual words used to invoke a given function may vary from system to system. The commands of CCL and their corresponding functions fall into four groups: session control, query formulation, output, and general. The CCL commands specified in the EURONET Guideline include the following:

Command	Function
Session control	
BASE	to select the data base to be searched
STOP	to end a session or part of it

Query formulation	
FIND	to enter a search statement
DISPLAY	to display a list of alphabetically or logically related search terms
SAVE	to save a search statement for later use
Output	
SHOW	to display or type records online
PRINT	to print records remotely (offline)
General	
DEFINE	to override default parameters and to create user macros
DELETE	to delete statements or print requests
MORE	to display more data
BACK	to display "previous" data
HELP	to obtain guidance online
NEWS	to obtain the latest information on the system
INFO	to give general information on aspects of the service (with specific subcommands EURONET, COST, SCHEDULE, USERS, STATUS)
OWN	to allow use of "original" commands

Certain general principles were followed in design of the language:

(1) all inputs are preceded by a command word unless in answer to a specific system request (a prompt);

(2) spaces are significant and must appear after each command word, operator, and identifier;

(3) two symbols are used as separators when specifying a list. When a comma (,) is used it implies that the list items are alternatives, and when a semicolon (;) is used the items must all be present;

(4) some commands will normally be followed by an identifier statement; where the statement entered lacks necessary details the system will respond by asking the user to supply any missing information.

The guideline also suggests standard two character labels for fields and data elements (e.g., abstract = AB, title = TI) and standard four character codes for language (e.g., English = ENGL, French = FREN). Negus and Snowden/[362] have prepared a CCL user's guide explaining its use for bibliographic reference retrieval. The guide is arranged to explain simple things before complex and also to represent the basic sequence of operations carried out in a typical search.

In practice there are actually three possibilities for the outcome of a CCL implementation by a given host: (1) the system provides all CCL functions--no more, no less; (2) the system provides some CCL functions and nothing more; (3) the system provides some or all of the CCL functions plus some additional facilities. Also in each of these cases the host may continue to support another interface in addition to CCL. The availability of additional facilities is recognized by the CCL command "OWN" which signals the system that a function beyond those provided by CCL is to be performed. The guideline recognized that each of the possibilities listed above was likely to occur because:

> (1) Many systems will have additional commands to enable functions not provided.
>
> (2) Systems are not expected to provide all the facilities available with each command, nor even necessarily all the commands. (Page 7)

Thus, while a CCL introduces some level of standardization across systems, it does not mean that retrieval systems implementing CCL become identical.

A number of Euronet hosts have already implemented versions of CCL: DIMDI, GID, INKA, FIZ-TECHNIK in Germany; ESA/IRS, CED, CNUCE, CILEA in Italy; ECHO in Luxembourg; and Télésystemès-Questel in France. CCL has been incorporated into or interfaced with several different retrieval packages, e.g., ESA-QUEST, GRIPS/DIRS, MISTRAL, STAIRS, ASPENSEARCH, 3RIP. Although implementation of CCL requires time and thus cost to accomplish, it is not a very difficult programming task.

Unfortunately the initial implementations of CCL by the various hosts were not without some inconsistencies. In particular Verheijen-Voogd/[511] presents a detailed comparison of the versions available through ESA/IRS and DIMDI. He found a number of variations in such things as the use of commas and semicolons and the use of dots as indications of patterns for word adjacency. At a meeting organized on 20 November 1981 in Luxembourg by the Euronet-DIANE Launch Team, representatives of the hosts that have implemented the CCL concluded that the small differences in the various versions should be ironed out as soon as possible. Thus there is some prospect that the variations identified by Verheijen-Voogd will be eliminated. Nevertheless, it is likely that some differences will remain because not all hosts will support all commands to the same extent. While the overall structure, syntax and semantics should be the same, some features, such as lefthand truncation and the ability to sort output, may not be provided by all hosts.

Implementations of CCL have not been restricted to Euronet hosts. The Information and Documentation Science Department of the Université Libre de Bruxelles has implemented a subset of CCL called SERIDU (Système Educatif de Recherche d'Information Documentation de l'ULB) (Van Slype et al/[508]). It is designed for use in-house as an educational tool for students. Larsson/[279] has described the implementation of CCL with 3RIP, a retrieval system developed in Sweden. While CCL was designed to cover searching only one data base at a time, 3RIP is designed to search several data bases in parallel. Thus some of the commands in CCL-3RIP do not strictly match those defined in the Euronet guideline. Some of the functions specified for the CCL commands BASE, FIND and STOP are transferred to 3RIP commands OPEN, CLOSE, RESTART and RUN in order to accomplish searching data bases in parallel. This is a good example of a system that provides some of the CCL functions plus some additional facilities.

Although CCL was originally designed to provide access to bibliographic data bases, Negus has more recently explored the feasibility of using it to search other types of data bases as well. His conclusions are mixed. With respect to terminology data banks, in particular EURODICAUTOM (the multilingual terminology data bank of the Commission of the European Communities), he has concluded that the retrieval facilities desired can all be expressed in principle using CCL (Negus/360). With respect to nonbibliographic retrieval systems in general, Negus/359 found too much diversity to make extension of CCL feasible. While CCL is capable of meeting most retrieval requirements of nonbibliographic retrieval systems, considerable extension would be needed to provide for the many data manipulation and sophisticated output features currently used in the various systems surveyed. The number of functions that would have to be provided is, in total, possibly too large to permit incorporation into a single language without introducing complicated conventions. Furthermore, many nonbibliographic retrieval systems employ a prompt-based rather than a command-based interface making a CCL even more unsuitable.

CCL illustrates standardization as a means of achieving compatibility among different retrieval systems. However, as can be seen from the preceding discussion, retrieval systems implementing CCL do not necessarily become identical. For the user for whom a few basic commands are sufficient to accomplish query formulation and output, essentially the same approach can be used on each host providing the CCL as a means of access. For the user who wishes to make use of more sophisticated features unique to a particular host, the option of using the "OWN" command is available. A number of advantages can be cited when CCL is widely implemented:

(1) users experienced in one or two systems can more easily search for information in related data bases on other hosts;
(2) users do not have to learn a new language every time they wish to search new systems and thus can concentrate on the search problem and any special data base features;
(3) users can be trained more easily and within a shorter period of time;
(4) new systems joining the network are more immediately accessible;
(5) the use of standard commands need not interfere with the provision of specialized search facilities on any system.

The acceptance of CCL as an international standard is likely to encourage even more widespread implementation.

As no command language is likely to remain static, it is necessary to provide some guidelines for the contolled development of CCL. Negus/359 has given the following recommendations as a means of gradually providing more and more facilities within the general framework of the language:

(1) Facilities should be considered for addition to the CCL when proposed by an actual or potential user.

(2) All proposed extensions should be subject to review before incorporation into CCL.

(3) All system operators using CCL should be given the opportunity to comment on proposed extensions and all views expressed should be given full consideration.

(4) A mechanism should be established to ensure that the latest version of CCL is made available - in the long term ISO might prove appropriate. (Page 34)

This case study illustrates the following section of the report: 4.2.

8.7 iNet: a case study in systems interconnection

iNet Gateway^{TM} is a national intelligent communications network, operated by the Computer Communications Group (CCG) of TransCanada Telephone System. Because the one year field trial began comparatively recently (July 12, 1982), it is not possible to present a detailed description of all features of iNet at this time. Nevertheless, iNet deserves to be the subject of a case study because it demonstrates an advanced approach to systems interconnection, with standardized procedures for the exchange of information among a variety of terminal devices and systems with different computer hardware and software. In particular it represents a coming together of videotex and online bibliographic data bases with easy modes of access by the end user (Nash/349). Tandem was selected as the supplier of gateway hardware and the software was created by Bell Northern Research.

The purpose of the one-year trial is to determine users' needs with respect to features, applications and interface requirements. User organizations have been classified into several industry or common interest groups: banking, bibliographic, communications, energy, government, real estate, legal and travel. The Bibliographic Interest Group includes the National Library of Canada, Canada Institute for Scientific and Technical Information, Université du Québec, University of Waterloo, University of Guelph, and Carleton University. iNet provides access to videotex, bibliographic, numeric and full text files offered by the following information service providers: B.C. Telephone, Bell Canada (Vista), Canada Systems Group, Canadian Press NEWSTEX, Concord Publishing, Infomart, Info Globe, Informatech, I. P. Sharp Associates Limited, NBTel, Official Airline Guides, QL Systems Limited, the Financial Post, and the Task Force on Service to the Public (Department of Supply & Services, Canada). Both the common interest groups and the information service providers are described in the first issue of the iNet Newsletter/238.

The trial will support up to 250 Telidon (videotex) terminals and 150 alphanumeric (ASCII) terminals. iNet is accessible from anywhere in Canada using Datapac, dedicated circuit service, or the public switched telephone network. Each user is recognized by the network as a unique individual through a discrete user profile. The profile defines how the user will interact with the system and includes specification of such parameters as terminal type (Telidon or ASCII), language (English or French), level of expertise (novice or expert), level of authority, and billing requirements. Consolidated billing is provided for the various services accessed.

Resources accessible through the network are identified through the use of online directories of three types: public, organization, and personal. The directories contain category listings and details on each service available. The public directory contains a consolidated listing of all publicly available services that can be accessed through iNet. An organization directory provides a complete list of all services available to a common interest group. Finally, a personal directory can be used to identify the services that an individual iNet subscriber calls on a frequent basis. Directories provide automated access to the listed information services and even automated "extended access" to particular data bases (when both an information service and a particular data base available through that information service are selected). The user signs on to the network only once per session. Thereafter access to services on the network is performed automatically when a directory item is chosen. No further identification is required since the network "intelligence" stores the user's profile. Both directories and user profiles are created and updated by CCG staff. All users also have access to the Envoy 100 messaging system which permits personal and group messaging, text editing, and multiple addressing.

By connecting a number of Telidon videotex hosts across Canada, including Infomart's Ottawa and Winnipeg computers, Bell Canada's Vista and hosts operated by B.C. Telephone and NBTel, iNet has created a national videotex network. In addition, Canada Systems Group, Canada's largest computer services company, is providing software that can convert standard data files to Telidon pages when the data are accessed. Thus a standard data base need not be restructured and stored in a specific Telidon data base for subsequent viewing. Instead, conversion is accomplished on demand when the file is accessed by a Telidon terminal.

The iNet Gateway™ responds to the growing demand for more universal access to information providers and computer-based services. Because the gateway has access to Datapac, it can even be used to link to U.S. and other international packet switched networks. Three key features provided by iNet are directories, ease of access, and personalization. Network "intelligence" performs many functions that users would normally have to perform themselves. The network stores the information needed to access various hosts and the user need not know all the variations in access protocols. In addition, centralized directories allow easy review of available resources followed by automatic linking to the one selected. The iNet Gateway™ provides an excellent example of how building "intelligence" into a network can compensate to a large extent for incompatibilities in both terminals and host hardware and software. Based on information gathered during the field trial, CCG plans to refine the iNet concept, conduct market research, and further develop an intelligent network responsive to the needs of users and information providers.

The applications being tested by the Bibliographic Interest Group during the one-year iNet Gateway™ trial suggest the ways in which such an intelligent network could support a decentralized nationwide bibliographic network (Durance et al[/130]). The participants have diverse hardware, software, and data base resources. The National Library of Canada's DOBIS system has a bilingual data base with two million records residing on an IBM computer. CISTI's CAN/OLE (CANadian Online Enquiry) system provides access to a wide variety of data bases via an IBM computer. Carleton University has its own automated system, CATSUP (CATalogue SUPport), allowing access to its library catalogue data base on a Honeywell mainframe. The University of Waterloo's library circulation system called CAM (Community Access Module) is housed on a GEAC minicomputer; the University of Guelph system called RAM (Remote Access Module) is similar to CAM. At the Université du Québec an online union catalogue is maintained for its intercampus library network. This French language data base called BADADUQ II (Banque de données à Accès Direct de l'Université du Québec) resides on a VAX minicomputer. The applications software for accessing and displaying data is unique to each system. Using iNet as a vehicle, the participating institutions seek to demonstrate the feasibility of an open network in which participants employ the network to facilitate interchange of information among incompatible terminal devices and computer hardware and software operated by autonomous institutions.

All Bibliographic Interest Group members are participating in the development and testing of network interfaces and directories in order that interconnection and access to participating bibliographic systems, as well as to other information provider data bases, can take place. In the past the Bibliographic Interest Group host systems have had to be accessed individually using logon procedures that are multistage, detailed, and prone to delay. The iNet trial seeks to overcome this by providing a listing of accessible systems, suitably categorized by type and allowing for automatic one-stage logon procedures to these systems.

A number of projects are planned to make use of resources through iNet. A data derivation project will assess the usefulness of utilizing different data bases for library cataloguing support operations by enabling users to access bibliographic records maintained on unilingual and bilingual systems. The electronic messaging capabilities of Envoy 100 will be used to assess the potential of new telecommunications services to improve interlibrary loan service effectiveness. Each member is creating a videotex data base and an analysis will be made of the strengths, weaknesses, and cost/benefits of videotex systems versus conventional bibliographic systems. Enhancements to videotex facilities to improve their usefulness for bibliographic applications will be identified. The network will also be used to assess the technical requirements for online file transfer of bibliographic data. The projects thus represent a wide range of activities related to library operations and services, and the experience with iNet should be suggestive of the benefits, pitfalls, and costs inherent in developing a nationwide decentralized bibliographic network.

This case study illustrates the following sections of the report: 5.6, 5.7.

8.8 CSIN: a case study in machine-assisted searching

CSIN, the Chemical Substances Information Network, is intended to offer the user an intermediary facility for accessing multiple online systems for related information. It assists the user in searching geographically dispersed, heterogeneous, online chemical information systems. Implemented initially as an intelligent terminal-based system that could be located at a user site, the CSIN prototype now runs on a VAX computer located at the offices of Computer Corporation of America (CCA) in Cambridge, Massachusetts.

Through the Toxic Substances Control Act of 1976 (TSCA), the United States Congress mandated the Council on Environmental Quality (CEQ) to study the feasibility of standardizing the classification, storage and retrieval of chemical information. Through the same act, the Environmental Protection Agency (EPA) was mandated to design and co-ordinate an effective system to reduce the information burden imposed by TSCA and similar legislation on groups in the public and private sectors. In 1978 the Interagency Toxic Substances Data Committee (ITSDC), a federal group cochaired by CEQ and EPA, was established to respond to these directives. The CSIN concept evolved to meet ITSDC's charge to seek ways to increase the efficiency and effectiveness by which organizations in both the private and public sectors can access and analyze chemical information. CCA is the development contractor for CSIN and Bolt, Beranek and Newman Inc. is the user support contractor for documentation, training, and assistance.

The development of CSIN involves three stages, which have been designated as preprototype, prototype version 1, and prototype version 2. They represent progressively more sophisticated interfaces and are upward compatible, i.e., the features of the preprototype are included in the prototype version 1 and the features of prototype version 1 are included in prototype version 2. The preprototype stage was also referred to as the CSIN Intelligent Terminal (CIT) because it was microprocessor-based. The prototypes, being more advanced and complex, have been implemented on a VAX minicomputer which can be reached through Telenet. The capabilities of each stage are listed below and an explanation of the various features follows.

Preprototype: Automation of Partial Queries

automatic connection to remote systems
data capture
data transformation
data substitution

Prototype Version 1: Automation of Complete Predefined Queries

scripts
user mail
+ preprototype capabilities

Prototype Version 2: Automation of Ad Hoc Queries

global query language
+ version 1 capabilities

CIT, the preprototype system, is particularly well suited for handling queries posed frequently, queries requiring minor changes to execute on another system, and queries using the results of previous queries to the same or different systems. Each function supported by CIT hardware and software has an associated benefit (Horowitz and Bergman[226]):

Function	Benefit
Connect: automatically conducts connection and login procedures for the user-selected system.	Relieves the user of the need to master the access procedure for each system.
Data capture: captures and stores locally all data coming from a remote system.	Provides storage of and convenient access to all retrieval data. Allows substitution of captured data in a query to another system.
Transform: automatically transforms data--input by the user or captured from a remote system--from the format of one data base to the format of another data base.	Greatly increases productivity by rapidly and reliably making transformations between different syntaxes.
Substitute: outputs data to a remote system as if it were typed in by the user.	Eliminates the need for the user to retype search terms for each data base. Thus, enhances reliability, encourages more comprehensive searching, and speeds up searches on multiple data bases.
Query lists: stores a list of search or index terms used to partially specify searches on a data base.	Increases the comprehensiveness of searches by allowing the creation of lists of index terms to specify the search for a particular topic. Decreases the chance of missing relevant information through not using all of a chemical's index terms.
Editor: Allows modifications of standard query lists or edits data obtained from one system before transforming it into a query on another system.	Provides flexibility by allowing the user to modify query lists of captured data to customize each query.

The usefulness of these functions can be illustrated by describing a typical sample search (Horowitz et al/227). Assume that a user wishes to gather all available online information about a chemical for which he knows the common name or the Chemical Abstracts Service registry number. The connection feature is used to access a chemical identification file such as CHEMLINE on MEDLARS or CHEMNAME on DIALOG and a query using the registry number or common name is transmitted. The data capture feature is invoked to store the output of the query--all synonyms and other identification information--on the terminal's diskette. The transform feature is used to reformat the captured data so that it can be used to query bibliographic files such as TOXLINE on MEDLARS. After connecting to the bibliographic file, the user instructs the terminal to substitute the reformatted data for keyboard input, thus fully querying the bibliographic file by using all known synonyms for the chemical. Additional systems can be accessed and queries made using the same techniques.

Given the experience gained with the preprototype CIT, CCA is marketing a commercial product CAST (Chemical Automated Search Terminal), described as "a system that increases productivity and effectiveness in multi-system searches for chemical information." The configuration of CAST includes a full duplex 1200 baud modem, a 64K microprocessor, two 500K floppy disk drives, a full ASCII printer, and a complete software system. CAST provides machine-assisted access to DIALOG, System Development Corporation's ORBIT, the National Library of Medicine's MEDLARS, and the NIH-EPA Chemical Information System. CAST can be used like a regular terminal to access other systems or additional custom interfaces can be provided by CCA.

The preprototype system can be thought of as providing "enhanced direct use" via a number of separate but related functions. The scripts of the version 1 prototype serve as a generalization and integration of these functions. The script is a predefined search on a specified subject which retains sufficient flexibility to permit modification each time it is used (Bergman/39). The user retains control of the search by selecting which data bases are accessed, what terms are used, and what the search strategy will be. Once query formulation has been completed, the script handles the resulting interaction with remote systems. The user thus needs no knowledge of remote systems in terms of command language, system protocol, or formats, as this is incorporated into the script. To initiate a search, the user is prompted via a menu-driven format to specify which data bases are to be searched, what search terms are to be used, and what search parameters are to be included as limitations. Once the user has selected the appropriate information, the script executes, intermittently informing the user what process is underway and displaying results from one data base before proceeding to the next. Examples of scripts include one on Chemical Identification which can access seven separate data bases and one on Toxic Health Effects which can access 23 data bases on five systems. Experienced searchers helped CCA with script definition by suggesting subjects to be searched, data bases to be included, and lists of search terms that would be useful. The script process can be represented in four steps:

1. Select script from user menu
   Environmental effects
   Health effects
   Chemical identification
   . . .
2. Select options or delimiters for search
   Date limitation
   Language
   Ad hoc query list
   . . .

3. Operation of script

4. Select output options
   View up to 15 citations
   Print
   File
   . . .

Currently scripts accept a Boolean expression of sets of search terms which are ANDed or ORed together as selected by the user. Other search parameters (such as language and date) are automatically ANDed together by the script. Predefined lists of terms making up stored query lists would be ORed together when sent to the remote system for searching. Examples of available query lists include Mammals and Carcinogenicity. Users can edit, create, and delete query list files as necessary to tailor a script to their needs. As the script executes, it transforms query lists into the correct format for each data base. The script notes and automatically handles special component features like the use of section codes in CA SEARCH. All scripts are programs written in a simple interpreted language. Eastlake et al/[134] present descriptions of two languages developed by CCA for CSIN: CSIL (the CSIN script implementation language) used for script writing and CPL (CSIN pattern language) used to define the patterns needed to accomplish data transformation between formats used by different data bases and/or systems.

Version 2 will go beyond version 1 by providing a global query language which will allow users to issue general requests for chemical information to the CSIN. Through distributed data base processing, the prototype would decompose the user's question into questions to appropriate data bases and systems, issue the component queries in the appropriate format and command language, retrieve the results, and integrate them into a response (Winter et al/[554]).

At present the prototype CSIN is being tested and evaluated by a number of user organizations of various types--corporations, federal agencies, state governments, and public interest groups. Resources available to users of CSIN include specialty files on MEDLARS, CAS-Online from Chemical Abstracts Service, NIH-EPA Chemical Information System, Chemicals in Commerce Information System (CICIS) maintained by EPA (an inventory of chemical substances manufactured, imported or processed in the United States), ORBIT files from System Development Corporation, and both the Chemical Regulations and Guidelines System (CRGS) and the TSCA Initial Inventory (a dictionary of chemical substances in commercial use in the United States) on DIALOG.

CSIN illustrates the use of machine assistance in dealing with incompatibilities found in multiple system searching. Without such assistance, the burden of dealing with the complexities of many uncoordinated systems is on the user. Normally the user is confronted with the following:

(1) lack of a common directory for all systems makes it difficult to know which data bases have information relevant to a search;

(2) many steps may be required to initiate communication with each separate data base;

(3) data bases may be available through different systems using incompatible command languages;

(4) lack of multisystem search co-ordination causes the user to repeat the same operations for every system searched.

In contrast CSIN provides the assistance needed for efficient and effective use of the resources of heterogeneous and geographically distributed systems. As described above, CSIN has the capabilities to:

(1) guide the user through the search process;
(2) determine where in the network information required by the user exists;
(3) formulate queries to various systems in the network to retrieve the required information;
(4) deliver a series of queries to appropriate systems and receive output from these systems;
(5) use results of queries (possibly modified by the user) in formulating queries to other systems;
(6) collect, sort, merge and deliver results to the user.

Rather than waiting for the various systems to make modifications that would lead to more compatibility (e.g., adoption of a common command language), CSIN can serve as an intermediary between the user and the various systems. As a node on Telenet, CSIN can be dialed up to initiate a search and it can in turn communicate with the various information systems which are Telenet hosts. The user is thus presented with an integrated view of diverse systems. In addition, the data capture, transformation, and substitution facilities can lead to reductions in time and cost required for searching.

The three stages of development of CSIN--preprototype and versions 1 and 2--represent increasing levels of machine assistance to the user. All provide augmentation, enabling the user to easily conduct searches that would require much time and knowledge of detail with conventional systems. Each stage involves more delegation of control to the system. In the preprototype stage the user formulates the search, using such facilities as data capture and transformation to automate selected steps. In the case of scripts, much of the strategy is already preprogrammed, but the user contributes to query formulation through such activities as entering terms, selecting data bases, and editing query lists. The global query facility of version 2, intended for "end users", shields the user from even more details of system operation. This approach is possible in the environment in which CSIN is intended to operate--a comparatively stable group of component information systems to be accessed and a well defined set of user data needs that can be supported by a limited class of query functions.

Future development of CSIN may lead to its implementation on a wide spectrum of computer systems, including, but not limited to, word processors, intelligent terminals, and personal computers (Lake/277). Powerful standard programming languages and operating systems are becoming increasingly common on microcomputers. This will make it possible to develop programs that can realize the CSIN concept, be maintainable, and portable.

This case study illustrates the following sections of the report: 4.4, 4.5.

9. CONCLUSIONS

One of the main objectives of the UNISIST programme as approved by the General Conference of Unesco in 1972 was "improving tools of systems interconnection". That this was listed first among the five main objectives is testimony to the importance assigned to the whole subject of compatibility within Unesco. The Division of the General Information Programme continues to be a catalyst in fostering compatibility among information systems and services throughout the world; in particular, by publishing various guidelines some of which (e.g., in thesaurus construction) have led to the development of ISO standards. Several of Unesco's activities in this area have been singled out for mention in this report. A compact account of these activities through 1975 can be found in Wysocki/561.

The importance of compatibility within the information systems of the United Nations itself was strongly re-emphasized in a recent report from a UN Expert Group/148. The Expert Group was critical of certain agencies of the UN for failing to adopt or adapt to existing standards:

> Many large UN family bibliographic data bases which establish focal points in the developing countries do not follow these standards. The result is that these agencies are exporting their own incompatibility to developing countries, which are trying to be compatible with the UNESCO standards. Our attention was also called to the need for co-ordination between several of the newly developed, or planned, large information systems in the UN family. (Page 10)

At the same time, however, the Expert Group is quick to point out that total compatibility, even within the UN itself, is likely to be unattainable:

> The whole exercise of co-ordination is not, in our view, meant to achieve total compatibility among all the information systems of the UN family of organizations. (Page 3)

and again:

> . . . throughout our interviews, perusal of records, readings, we were never met by any call for, or by an indication by anybody of, the desirability of total compatibility among information systems. (Page 15)

While recognizing the importance of co-ordination, the Expert Group also expressed awareness of the difficulties:

> Nobody has said that co-ordination can be achieved by dictat, or that it can be achieved quickly. It is a process which takes time and involves the rubbing down of many rough edges, but it has to take place. If arguments are presented to show how impossible the task is and how futile efforts are, then such arguments can be considered nothing more than an excuse and need to be recognized as such. (Page 16)

According to another recent UN report/500, the first task assigned to the Pan African Documentation and Information System (PADIS) was "to standardise documentation procedures at the operational level in Africa." The report goes on to complain that, despite Unesco's efforts, "even the most active UN agencies in

Africa (like UNESCO and FAO) worked according to different formats and procedures for document analysis."

Throughout the present report, the authors have striven to identify various types and levels of compatibility affecting the operation of information services. In this, they have allowed compatibility to, as it were, "speak for itself". That is, they have tried to illustrate how compatibility can serve to facilitate resource sharing, improve efficiency and reduce costs of some aspects of information service. The message seems clear: information centres cannot afford to be completely self-sufficient; adherence to well-established standards greatly improves the probability that one organization can make effective and economical use of the products and services of others. Emerging information services in the developing countries have more to gain than most from the adoption of appropriate international standards: they will be able to co-operate more effectively with each other to conserve their limited resources and will find it easier to exploit the richer resources of the mature information services of the developed world.

The importance of compatibility to the information services of the developing countries has been presented very cogently by Wild and Woolston/543:

> Developing countries, simply because they may have invested less in information files than have developed countries, are often better able to adopt standards than are developed countries. And if the developing countries will make the small investments needed now to adopt international standards in building their national information services, they will avoid the much higher costs that would be incurred in the future in readjusting their methods to achieve compatibility with those of the world community. (Page 7)

And again:

> Developing countries will not all be able to afford the luxury of establishing new national information structures to facilitate participation in each new international information system. They will often need to centralize their activities within one unit in order to capitalize on trained manpower and equipment. UNISIST, working towards a family of compatible information systems, facilitates such centralization. INIS and AGRIS (and perhaps DEVSIS too) can share a central processing facility because the technical specifications of the two systems are very similar. This also means that people trained to do bibliographic cataloguing in national centres could catalogue INIS as well as AGRIS literature. The same people could be trained to produce paper-tape or magnetic-tape files for input to the two systems; and the same techniques for providing individualized services to users at the national level could be applied.
>
> If international information systems share common central processing facilities, this is a financial benefit to all countries that contribute to the budgets of the international organizations concerned, because the equipment (computers, microfiche equipment, etc.) and the training of staff do not have to be duplicated, and because the utilization of this equipment and staff benefits from the economy of scale. The countries also gain the advantages of common international training programs; and their nationals, who do receive

> such international training, can spread what they have learned at the national level with the assurance that it will be relevant in more than one narrow environment. (Pages 15-16)

It would be misleading for us to imply that standards provide a universal panacea for information services. In actual fact, while standardization can resolve various problems, it creates difficulties of its own.

In the fields of library science and documentation, standards tend to derive from the practice of organizations whose activities are accepted as being, in some sense, a model to be followed. Thus, to mention a single, obvious example, ISO 2709 derives directly from the experience with MARC. Clearly, since standards result from operating experience with information services, their publication will not be "in phase" with the needs of many agencies. These agencies, unable to wait for a standard to emerge, may have committed themselves to practices and protocols that are not fully in consonance with the standard when it finally appears. To convert to the standard, especially if this involves retrospective processing, may be an expensive proposition. Even within the United Nations, complete fidelity to international standards has been difficult. The extent to which one UN system, INIS, has been able to adhere to international standards is fully discussed by Gove et al/[187]. When AGRIS was planned, the UNISIST *Reference Manual* did not exist, necessitating the development by FAO of its own *AGRIS Guidelines for Bibliographic Description* (East and Martinelli/[133]). In this case, the standard was, as far as possible, made compatible with the experience of existing information services, including AGRIS, rather than the reverse. This particular difference in phasing has created problems in the implementation of other UN systems, notably in the attempt to make DEVSIS compatible with the RM as well as with INIS and AGRIS. As Morin-Labatut and Sly/[338] have pointed out:

> The differences are not negligible. However, a series of compromises have been proposed. With AGRIS as an already operating system with more than 720,000 records in its data base, the flexibility has had to come mainly from the other side.

A somewhat similar situation has existed vis-a-vis the MARC format. To give one illustration, when the National Library of Medicine developed its MEDLARS system, the MARC format did not exist. Moreover, the format had not been finalized when NLM was developing its second-generation system (MEDLARS II). Only in its current development effort (MEDLARS III) has the Library been able to achieve a higher level of compatibility with MARC.

The fact is that computer and telecommunications technologies are moving too fast for the standards bodies. Progress cannot be held up while appropriate standards are developed. Standards derive from operating experience. There is a pragmatic implementation of standards within information systems, particularly those that are most innovative, long before the standardizing agencies become directly involved.

Another problem is that certain types of international standards are not sufficiently complete for various information service applications. In the United States, for example, the National Library of Medicine needs to uniquely identify every serial publication appearing in its MEDLINE data base. Unfortunately, ISSNs do not exist for all titles processed by the Library. For new titles, the ISSN may not be assigned until after the Library has been handling the title for some months. Although, where available, ISSNs are carried in MEDLINE records, the fact that they do not always exist makes them unsuitable for use as internal control numbers in this and related data bases. Consequently, the Library has developed its own unique set of "journal title codes".

The same limitation that applies to the ISSN applies to the CODEN of the American National Standards Institute--the codes are not available for all serials or not available fast enough. In fact, the Chemical Abstracts Service adopted official responsibility for developing new CODEN for this very reason: to have the codes available for new serial publications as soon as CAS needs them.

Stokely/468 has suggested that the major advantages to the adoption of standards in information systems are ease of system interconnection, possible cost savings (e.g., use of standard software), and possible improvements in quality of service (if standards serve to raise levels in some centres). Yet, adoption of national or international standards may involve some conversion costs and may force an agency into making further costly changes as the standards themselves are updated. Thus, some loss of stability, as well as loss of autonomy, may occur. On the other hand, Middleton et al/333 stress the importance of standards in achieving any useful level of interconnection among information systems.

While dealing specifically with the difficulties involved in developing a common command language (CCL) for interrogating numeric data bases, Negus/359 gives a concise summary of some of the major pros and cons relating to standardization in general:

> While at a theoretical level a strong case can be made for standardisation of the user-interface, in practice there are number of objections which can be put forward.
>
> The case for standardisation is based on a number of arguments:
>
> - Many functions are common to a majority of systems.
>
> - Use of a standard interface means users can readily move from system to system, and use new systems. They are consequently able to concentrate on obtaining best results from the data and facilities available on each system.
>
> - Standardisation aids the system operator who is able to change software and/or hardware without worrying about retraining of users and any consequent decrease in usage.
>
> Objections to the adoption of standardisation are really of two types. Firstly there is the theoretical argument which comes in several different guises all of which have the basic premise that standardisation is bad because it will inhibit the development of new facilities. This limitation, which is anyway not true if standards are developed with awareness of the danger, is seen as stifling innovation, as being likely to force system operators to use one of an ever smaller number of hardware/software combinations, and as preventing competition. With reasoned argument all these objections can be shown to be untrue in general terms; indeed, development and widespread adoption of good standards can actually stimulate innovation and give more scope for competition between hardware, between software and between service operators.
>
> The second type of arguments are of a practical or technical nature and more weight must be attached to these. At the heart

of the arguments lies the belief, which is not without foundation, that while standardisation is desirable in theory it is unlikely to be feasible, certainly in the short to medium term and possibly never, because:

- The total range of facilities already provided is enormous
- Each system has its own individual requirements
- Many different database structures are in use
- Some systems will never be used outside a relatively small group of users
- A generalised interface would be less attractive to the specialist user than one developed for limited precise purposes
- Some systems make use of specialised output devices, e.g., graphic terminals, which can be provided with special function keys to make use simple and efficient
- The amount of software development required to provide a new interface may often be prohibitive.

Wellisch/531 claims three major types of problems associated with standardization:

> The generation and display of bibliographic data in databases as well as in conventional printed publications suffers at present from a complete lack of uniformity which is due to three different conditions of standardization:
>
> (a) the existence of a multiplicity of standards which are either overlapping, mutually contradictory, or applied in sometimes quite different ways by the producers of recorded information;
>
> (b) a lack of standards, resulting in idiosyncratic rules and solutions of certain problems; and
>
> (c) the persistence of obsolete quasi-standards, invented by unknown people on their own authority, sometimes in a distant past, when they were geared to publication methods and patterns that have now themselves become entirely obsolete. (Page 213)

While he was specifically referring to consistency in bibliographic citation, Wellisch puts his finger on a situation that is probably true for information services in general: for some aspects, competing or overlapping standards exist; for others, standards are essentially lacking; for yet others, the standards followed derive from earlier practice that may have little relevance to the existing situation. Numerous examples of this last category seem to exist. For example, some standards relevant to card catalogues may have little applicability to computerized catalogues, while some levels of compatibility desirable in offline computer systems may not have the same degree of importance in online systems where, through direct interaction with a data base, a user can compensate more easily for certain types of variability in the presentation of data. Perhaps the most obvious case, however, is the difference between the requirements for producing printed products, such as indexes, and for producing equivalent products in machine-readable form only. For instance, a controlled vocabulary may have much greater applicability to the printed tool than it does to a machine-readable data base that can be searched online.

Park/382 gives some support to Wellisch's criticisms in suggesting that some standards in the field appear to be developed in a rather haphazard manner:

> With the notable exception of library cataloging, the journal literature contains very little information about standardization in the bibliographic field. Neither the issues nor the solutions are identified and argued publicly; rather, the deliberations tend to be confined to small, volunteer committees or working groups, dedicated to the proposition that standards are a necessary precursor to effective cooperation, especially in the evolving computer-based network environment. If standards are, indeed, as necessary as is generally believed, then this argues for increased visibility, staffing, and financial support appropriate to the level of contribution made toward national and international resource sharing. It has also been noted that much of the current standards development is based on "fingers stuck in the wind," as far as detailed study of the purposes and underlying fundamentals are concerned. This, too, probably reflects, in part, the inadequate base of management and research and development support for the standardization activities per se. One of the most important concepts coming out of the standardization work associated with the data base management approach to data processing may be the data base administrator as a focal point and cost center associated with information systems.
>
> A comprehensive framework, or plan, for standards development appears to be lacking. It is not unusual to find the same standardization activities being pursued by several groups simultaneously. Even worse, the order in which problems are addressed is often wrong, with the result that standards essential to continued progress are not available or are inappropriate to subsequent levels of work. While an evolutionary, rather than a revolutionary, approach to standardization is probably essential to maintain continuity of bibliographic services, it is still necessary to plan this evolution in an orderly fashion that optimizes scarce resources and builds on prior work without unnecessary duplication of effort. Organizations which will be faced with implementation of national and international standards need to be aware of the total picture so they can plan for orderly implementation of individual standards as they are promulgated; they require some confidence that their investments will not be obviated immediately as a result of substantive revisions and/or other standardization activities. There is also a need for investigating different kinds of standards--ones which are elegant in their simplicity and extensibility, like those of the ISO ASCII character set.
>
> To end on a more optimistic note, the small group of specialists who have dedicated long hours of time--often, personal time--to bibliographic standards activities are to be commended for the progress that has been made over the past decade. Without this volunteered effort, there would have been no standards at all to

> review for ARIST. Now is the time to look toward moving the standardization process, and its products, out of "art" and more into the "sciences," making greater use of planning tools and methodologies which have proven useful in other areas of cooperative endeavor. (Pages 70-71)

Referring specifically to the standardization of data elements in numeric data bases, Nolan and Whitney/372 show when and why standards are needed, as well as giving some reasons why their use may be resisted:

> Perhaps in a technologically ideal world, standardization would be unnecessary. All Fords were black only until the technology of assembly line management permitted red car bodies to be joined to red car doors and sent to a customer who wanted a red car. Data base standards would be unnecessary if computers could automatically and efficiently translate files into whatever format was required by the immediate user.
>
> Standardization is an administrative technique for coping with diversity. It reduces the opportunities for confusion and it facilitates the replication of results. Standards are generally introduced when a human activity has grown in size to the point where it is no longer controllable on an *ad hoc* basis. Applied to data this means that by limiting the data system to a standard set of conditions, the problems of controlling the data system are reduced to a manageable scope. Because a computer's operating system cannot automatically handle the diversity of data base definitions and structures envisioned by system designers, and since users too are having trouble dealing with comparable but incompatible data sets, it is necessary to restrict the designers of data files to a standard structure and standard definitions. But in doing this we should note that because standards are usually a constraint on the designer's freedom, they are likely to be met with resistance. Therefore, if one wishes to have proposed standards accepted, one must show how those standards *improve* the functioning of the *overall* data management and analysis process. In particular, in proposing standards it is necessary to anticipate their impact on forms designers, file designers and data users.
>
> Justifying new standards is always difficult, but the criteria are clear. For the most part, the costs are tangible and immediate while the benefits are intangible and distant. Often, benefits accrue to one group while another group pays the costs. (Page 4)

The biggest problem to be guarded against by an information service, however, is a possible conflict between the requirements of compatibility and the needs of users. The more specialized the service, perhaps, the less likely it is that the needs of the users will adapt well to some standard treatment. As pointed out earlier, "specialization" is virtually the antithesis of "standardization." An information centre must be careful not to dilute its effectiveness for the sake of achieving compatibility with other institutions. To take a hypothetical case, it would make little sense for an information centre in the field of irrigation to adopt a general agriculture thesaurus to achieve compatibility with other centres. In all probability, the general thesaurus would lack the detailed irrigation terminology needed to meet the needs of users. On the other hand, it might make a lot of sense to construct an irrigation microthesaurus that is fully compatible with the more general tool. An important lesson can be learned from this example: in a particular situation one means of achieving compatibility may be quite acceptable while another is not.

Heymans/215 is one of several authors who have claimed that standards satisfying the needs of computers may be quite inappropriate to the needs of humans. He claims, for example, that the punctuation conventions of the ISBDs, even if relatively unambiguous in machine processing, may be far from obvious to the human reader. As did Wellisch/531, Heymans also suggests that certain "standard" practices derive from long traditions but lack strong empirical support. He instances the preference given in cataloguing codes to author access points despite the fact that numerous studies have consistently shown that library users are more likely to have complete and correct title information than they are to have the name of an author in complete and correct form.

Mackenzie Owen/297 is even more insistent that the needs of users must always be uppermost in mind. Standardization activities that assume the existence of "one ideal description of a document", he claims, fail to account for the "inevitable differences between various groups of users." In a perfect information system (perfect from the viewpoint of "user satisfaction"), there should be as many descriptions of a document as there are differing groups of user needs. Standardization improves the efficiency of information services but diversification, based on the recognition of differing needs, promotes effectiveness. Mackenzie Owen believes that some acceptable middle ground must exist. Above all, any approach to achieving compatibility must be flexible enough to accommodate differing local requirements.

Throughout this report, two alternative approaches to achieving compatibility have appeared repeatedly: standardization and conversion. Adoption of standards (whether for hardware, software, document representation, or index languages) can lead to more compatible information systems but may introduce some inflexibility, the inability to tailor a system to local needs. New developments in technology, as described in the sections on hardware and software, have made conversion an alternative means of achieving compatibility in many cases. Where no standards exist or where standards are felt to be too restrictive to satisfy local needs, mechanisms for conversion may still allow a given system to achieve compatibility with another system for purposes of sharing resources. Although in principle convertibility has been recognized as an approach to compatibility for some time, in practice it is only with new developments in technology that it has emerged as a feasible alternative to standardization for reducing the many sources of incompatibility among systems.

In considering the factors affecting information system compatibility, it is clear that hardware and software differ from document representation and index languages in terms of the likely sources of developments that would lead to greater information system compatibility. While initiatives for standardization of hardware and software may come from a wide variety of application areas, the responsibility for promoting increased standardization in aspects of design specific to information systems (e.g., vocabulary compatibility) lies with developers of information systems. Thus, information system designers and users can benefit from developments in hardware and software compatibility which originate in other areas, but they themselves must assume responsibility for initiatives to achieve compatibility at the levels of information organization and representation once the technological infrastructure is in place.

In the main sections of this report various problems were identified with the underlying implication that these represent the areas in which more research or standardization needs to be applied. It seems pointless to repeat these specific and implicit recommendations here. Instead, we will conclude with some general observations that may guide Unesco and other organizations as to future priorities for standardization and normalization activities.

At the beginning of this report a general information transfer cycle (Figure 1) was briefly discussed. When this cycle is examined in relation to recent achievements in the area of compatibility, a number of conclusions seem inescapable. First, much of the compatibility effort has been concentrated on the later stages of the cycle, especially the operations of libraries and other types of information services. Much less compatibility appears to exist at earlier stages in the cycle--those pertaining to primary and secondary publication. Moreover, little compatibility extends across the various phases of the information transfer cycle. To mention an obvious example, standardization of bibliographic references by primary publishers could substantially facilitate the activities of other elements in the cycle: those performed by authors, secondary publishers, libraries and information centres of all types.

A more global approach to compatibility issues, taking into account potential interrelationships among the various information transfer agencies, seems highly desirable. In particular, national and international agencies concerned with standardization should give more attention to the earlier stages of the cycle since, in the long run, the earlier the standardization occurs the more cost-effective it is likely to be. If more standardization could be achieved in primary publishing--in bibliographic referencing, in the adoption of a limited set of "standard" type faces (to facilitate subsequent optical character recognition), and so on--many later operations could be performed more efficiently. Carried to its logical conclusion, of course, standard bibliographic records in machine-readable form would be generated by (for example) book publishers or, alternatively, these publishers would include OCR-readable standard bibliographic descriptions (ISBDs in effect) on the reverse of each title page.

A closely related conclusion is that certain standardization activities could achieve much more if they were carried one stage further. For example, the ISSN is extremely useful as a unique identifier of periodicals and other serial publications. How much more valuable it would be if extended to an international standard article number and how comparatively easy to implement: beginning, say, in January 1984 a publisher of a journal or series would simply assign a consecutive number to each article published, the ISSN followed by a unique article number. Had this simple device been adopted when the ISSN was implemented, several existing problems could have been avoided. In particular, when machine-readable data bases, or search results from these, are merged, duplication could easily be eliminated. Moreover, a unique article identifier of this type could also facilitate the payment of royalties to publishers, the ordering of photocopies, and other activities of concern to publishers, libraries and other agencies. Fortunately, the "add-on number" proposed to ISO by Ehlers/[138] could, if adopted, serve these various purposes.

In conclusion, the authors strongly urge that Unesco, IFLA, ISO, ICSU, and other international agencies, focus more attention on compatibility issues affecting multiple stages of the information transfer cycle. Moreover, this attention should extend beyond the present situation. It now seems very likely that various forms of electronic communication will supplant print on paper in many facets of human communication. It is possible to conceive of a system in which a scholarly article is created in machine-readable form by the author, accepted into a machine-readable data base (the equivalent of a printed journal) in essentially the same form, accessed online in that form, possibly read off and stored in many "personal" data bases, printed out at various times for various purposes, and so on. In such an electronic communication system, many existing distinctions--in the operations of primary and secondary publishers, libraries, and information services of all kinds--will become increasingly blurred. As this occurs, a more integrated, global approach to compatibility issues will become increasingly possible and increasingly necessary. Since fully electronic communication systems are merely in an embryonic stage of development, a unique opportunity exists for avoiding, ab initio, much of the incompatibility associated with the present print-on-paper environment.

REFERENCES CITED

1. ADLER, E. Hebrew cataloging and the computer--the view from Israel. *Information technology and libraries* (Chicago), Vol. 1, No. 3, 1982, pp. 238-245.

2. AGENBROAD, J. E. Character sets: current status and East Asian prospects. *Journal of library automation* (Chicago), Vol. 13, No. 1, 1980, pp. 18-35.

3. AHUJA, V. *Design and analysis of computer communication networks*. New York, McGraw-Hill, 1982, XIV, 306p.

4. AILAMAZYAN, A. K.; VESELOVSKII, A. V. ASIST-2: software and hardware for telecommunications. *Automatic documentation and mathematical linguistics* (New York), Vol. 14, No. 2, 1980, pp. 1-5. [English translation of *Nauchno-tekhnicheskaya informatsiya, seriya 2* (Moscow), Vol. 14, No. 3, 1980, pp. 1-4.]

5. *American National Standard for bibliographic references*. New York, American National Standards Institute, 1977, 92p. (Z39.29-1977)

6. ANDERSON, D. Bibliographic control in practice. *IFLA journal* (Munich), Vol. 6, No. 3, 1980, pp. 242-248.

7. ANDERSON, D. IFLA's programme of ISBDs. *UNESCO bulletin for libraries* (Paris), Vol. 32, No. 3, 1978, pp. 144-150.

8. ANDERSON, D. International Standard Bibliographic Descriptions in non-roman scripts. *Unesco journal of information science, librarianship and archives administration* (Paris), Vol. 3, No. 3, 1981, pp. 178-181.

9. ANDERSON, D. *Universal bibliographic control: a long term policy. A Plan for action*. Munich, Verlag Dokumentation, 1974, 87p.

10. ANDERSON, G. A.; JENSEN, E. D. Computer interconnection structures: taxonomy, characteristics and examples. *ACM computing surveys* (New York), Vol. 7, No. 4, 1975, pp. 197-213.

11. ANDRE, P. Q. J. Library of Congress automated authority systems. *In*: FURUYA, N. Y., *ed*. *What's in a name? Control of catalogue records through automated authority files*. Toronto, University of Toronto Library Automation System, 1978, pp. 31-54.

12. ANSART, J. P. Development and certification of standardized protocols. *In*: CSABA, L.; SZENTIVÁNYI, T.; TARNAY, K., *eds*. *Networks from the user's point of view* (proceedings of the IFIP TC-6 working conference COMNET '81, Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 477-487.

13. ANTOPOL' SKII, A. B.; RUDYKINA, N. G. Opyt sravnitel' nogo issledovaniya otechestvennykh informatsionno-poiskovykh tezaurusov. *Nauchnye i tekhnicheskie biblioteki SSSR* (Moscow), Vol. 12, 1977, pp. 3-11.

14. ARMSTRONG, C. *Software for printed indexes: a guide*. London, Aslib, 1982, 98p.

15. ATHERTON, P. Standards for a user-system interface language in on-line retrieval systems: the challenge and the responsibility. Online review (Oxford), Vol. 2, No. 1, 1978, pp. 57-61.

16. AUBER, P. R. My wish is your command?? In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 433-445.

17. AUBRAC, R. AGRIS--problèmes de compatabilité et d'interconnection. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 35-40.

18. AVRAM, H. D. Network-level decisions: basis and key issues. In: KENT, A.; GALVIN, T. J., eds. Information technology: Critical choices for library decision-makers. New York, Marcel Dekker, 1982, pp. 157-167.

19. AVRAM, H. D. Recommendations towards the establishment of a common exchange format for use by all agencies in the information community. In: DIERICKX, H.; HOPKINSON, A., eds. Toward a common exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 135-146.

20. AVRAM, H. D.; MCCALLUM, S. H. Electronics and cataloguing, twenty years later. International cataloguing (London), Vol. 10, No. 1, 1981, pp. 8-10.

21. BADRAN, O. A.; HAMAN, J.; LANCASTER, F. W.; MARTYN, J. Report on the independent appraisal of AGRIS. Paris, Unesco, 1977, VI, 118p.

22. BAKER, D. B. Harmonization of national policies to improve interconnection and compatibility. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 247-255.

23. BAKONYI, P.; KISS, I.; PETRENKO, A.; SEBESTYEN, I. Promotion of East-West computer communication in IIASA's international environment and the Hungarian case study. In: CSABA, L; SZENTIVÁNYI, T.; TARNAY, K., eds. Networks from the user's point of view (proceedings of the IFIP TC-6 working conference COMNET '81, Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 119-125.

24. BARCHANSKI, J. Z. Some aspects of network user access procedures design. In: CSABA, L.; SZENTIVÁNYI, T.; TARNAY, K., eds. Networks from the user's point of view (proceedings of the IFIP TC-6 working conference COMNET '81, Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 547-557.

25. BARRENTINE, J. K. Some technological issues of a national library network. Bulletin of the American Society for Information Science (Washington, DC), Vol. 5, No. 5, 1979, pp. 29-30.

26. BARRETT, R. Developments in optical disk technology and the implications for information storage and retrieval. Journal of micrographics (Silver Spring, MD), Vol. 15, No. 1, 1982, pp. 22-26.

27. BARRETT, R. Prospects for facsimile in information transfer. In: The impact of future developments in communications, information technology and national policies on the work of the aerospace information specialist. Neuilly sur Seine, North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development, 1977, pp. 7-1 - 7-10. (AGARD conference proceedings No. 225)

28. BARTHOLOMEW, M.; GEBREMEDHIN, E. The PCM vendors. Datamation (New York), Vol. 25, No. 2, 1979, pp. 104-111.

29. BARTLETT, K.; RAYNER, D. The certification of data communication protocols. In: Trends & applications: computer network protocols. New York, Institute of Electrical and Electronics Engineers, 1980, pp. 12-17.

30. BATTELLE COLUMBUS LABORATORIES. Linking the bibliographic utilities: benefits and costs. Columbus, Ohio, 1980, 169p. plus appendices.

31. BAUER, G. Zur methodik des aufbaus "koordinierfähiger" fachthesauri im rahmen des thesaurussystems chemie. In: Thesauri für mechanisierte Informations recherchesysteme. Berlin, ZIID, 1967, pp. 207-209.

32. BEHRENS, H.; EBEL, G. Numerical data banks in physics: what has been done, what should be done. In: GLAESER, P. S., ed. Data for science and technology: Proceedings of the Seventh International CODATA Conference, Kyoto, Japan, 8-11 October 1980. Oxford, Pergamon Press, 1981, pp. 212-218.

33. BELING, G. The use of EDP in terminological work. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 101-121.

34. BELING, G.; WERSIG, G. The new concept of an intermediary language system for information networks. In: BATTEN, W. E., ed. EURIM II: a European conference on the application of research in information services and libraries. London, Aslib, 1977, pp. 117-121.

35. BELLONI, A.; LE MOLI, G. Protocols standardization versus protocols conversion in computer networks interconnection. In: BEAUCHAMP, K. G., ed. Interlinking of computer networks; proceedings of the NATO Advanced Study Institute held at Bonas, France, 28 August - 8 September 1978. Dordrecht, D. Reidel, 1979, pp. 111-150.

36. BERG, J. L., ed. Data base directions--the conversion problem. Washington, DC, National Bureau of Standards, 1980, 167p. (Spec. Publ. 500-64)

37. BERGER, M.; KLEMPERER, K. Developing an effective man/machine interface for a large on-line university library catalog. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 18, 1981, pp. 267-269.

38. BERGER, P. W.; TUCKER, J. C. Standards and guidelines for data. In: ROSSMASSLER, S. A.; WATSON, D. G., eds. Data handling for science and technology: an overview and sourcebook. Amsterdam, North-Holland, 1980, pp. 93-113.

39. BERGMAN, R. F. Beyond SDI: on-line retrieval scripts in the Chemical Substances Information Network. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 18, 1981, pp. 276-278.

40. Bibliographic guide for editors and authors. Washington, DC, American Chemical Society, 1974, 362p.

41. Bibliography of United Nations thesauri, classifications, nomenclatures. Geneva, Inter-Organization Board for Information Systems, 1979, VIII, 252p.

42. BIRNBAUM, J. S. Computers: a survey of trends and limitations. Science (Washington, DC), Vol. 215, No. 4534, 1982, pp. 760-765.

43. BLAIR, J. C., JR. Micros, minis and mainframes...a newcomer's guide to the world of computers--especially micros. Online (Weston, CT), Vol. 6, No. 1, 1982, pp. 14-26.

44. BOCHMANN, G.; GECSEI, J. Towards videotex standards. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 253-262.

45. BOIKO, N. V.; POLOZHENTSEVA, G. I. Compatibility of formats for representing bibliographic information in automated IRS's. Automatic documentation and mathematical linguistics (New York), Vol. 14, No. 3, 1980, pp. 31-40. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 14, No. 5, 1980, pp. 17-25.]

46. BOLLMANN, P.; KONRAD, E. Automatic association methods in the construction of interlingual thesauri. In: BATTEN, W. E., ed. EURIM II: A European conference on the application of research in information services and libraries. London, Aslib, 1977, pp. 152-155.

47. BOUILLÉ, F. Data portability between large and mini-computers. In: GLAESER, P. S., ed. Data for science and technology: Proceedings of the Seventh International CODATA conference, Kyoto, Japan, 8-11 October 1980. Oxford, Pergamon Press, 1981, pp. 508-511.

48. BOURNE, C. P. Frequency and impact of spelling errors in bibliographic data bases. Information processing and management (Oxford), Vol. 13, No. 1, 1977, pp. 1-12.

49. BOYCE, C. MINISIS. Program (London), Vol. 16, No. 3, 1982, pp. 131-141.

50. BOYLE, S. O.; MILLER, A. P. Feature comparison of an in-house information retrieval system with a commercial search service. Journal of the American Society for Information Science (New York), Vol. 31, No. 5, 1980, pp. 309-317.

51. BRANSCOMB, L. M. Electronics and computers: an overview. Science (Washington, DC), Vol. 215, No. 4534, 1982, pp. 755-760.

52. BRIGHT, R. D. Videotex and the French telematique programme. In: 4th international online information meeting, London, 9-11 December 1980. Oxford, Learned Information, 1980, pp. 63-66.

53. BRITISH COMPUTER SOCIETY. Query Language Group. Query languages: a unified approach. London, Heyden & Son, 1981, 105p.

54. BROOKS, K. M. The online transfer of machine-readable data: a Pandora's box. Database (Weston, CT), Vol. 5, No. 1, 1982, pp. 18-21.

55. BROWN, P. J. Portability. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 1085-1086.

56. BROWN, P. J., ed. Software portability: an advanced course. Cambridge, Cambridge University Press, 1977, XIV, 328p.

57. BROWN, T. P. Communication standards for online interchange of library information. Library trends (Champaign, IL). Vol. 31, No. 2, 1982, pp. 251-263.

58. BROWN, T. P.; DEBUSE, R. Replicating the Washington Library Network computer system software. Journal of library automation (Chicago), Vol. 14, No. 3, 1981, pp. 202-205.

59. BROWNRIGG, E. B.; LYNCH, C. A. Packet switching and library automation: a management perspective. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 19, 1982, pp. 54-57.

60. BSI ROOT Thesaurus. Hemel Hempstead, British Standards Institution, 1981.

61. BSO--Broad System of Ordering: schedule and index. The Hague, International Federation for Documentation 1978, XIV, 184p.

62. BUCHINSKI, E. J.; ISLAM, M. The context of interconnection for a nationwide bibliographic network. Ottawa, National Library of Canada, 1980, VII, 33p. (Canadian network papers No. 1)

63. BUFFET, P. Exchange of bibliographic data: problems raised by the carrier. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 21-24.

64. BUNCH, S. R.; ALSBERG, P. A. Computer communication networks. Annual review of information science and technology (White Plains, NY), Vol. 12, 1977, pp. 183-216.

65. BURNS, C. Information storage and display. Journal of the American Society for Information Science (New York), Vol. 32, No. 2, 1981, pp. 141-147.

66. BUTLER, M.; BRANDHORST, T. Construction of a tertiary data base: the case of the Women's Educational Equity Communications Network. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 17, 1980, pp. 174-176.

67. BYKOVA, L. N.; MUZRUKOV, N. B. Experience gained in establishing a bilingual thesaurus for indexing documents relating to atomic science and technology. In: Handling of nuclear information. Vienna, International Atomic Energy Agency, 1970, pp. 545-555.

68. CADWALLADER, G. Format compatibility and conversion among bibliographic data bases. Philadelphia, Auerbach Corporation, 1969, 16p.

69. CAHN, D. F.; HERR, J. J. Automatic database mapping and translation methods. Berkeley, University of California, Lawrence Berkeley Laboratory, 1978, 13p. (LBL-6782)

70. CALK, J. On-line authority control in the Washington Library Network. In: FURUYA, N. Y., ed. What's in a name? Control of catalogue records through automated authority files. Toronto, University of Toronto Library Automation System, 1978, pp. 135-159.

71. CALLAHAN, M. V.; RUSCH, P. F. Online implementation of the CA SEARCH file and the CAS registry nomenclature file. Online review (Oxford), Vol. 5, No. 5, 1981, pp. 377-393.

72. CASEY, M. Euronet-DIANE: a study on the harmonisation of user manuals. Online review (Oxford), Vol. 4, No. 1, 1980, pp. 33-40.

73. CASEY, M. Packet switched data networks: an international review. Information technology: research and development (Guildford), Vol. 1, No. 3, 1982, pp. 217-244.

74. CHAN, L. M. AACR 6 and the corporate mystique. Library resources and technical services (Chicago), Vol. 21, No. 1, 1977, pp. 58-67.

75. CHAREN, T.; GILLESPIE, C. J. MEDLARS abbreviations for medical journal titles. Bulletin of the Medical Library Association (Baltimore), Vol. 59, No. 3, 1971, pp. 420-429.

76. CHERNYI, A. I. Information retrieval languages: types, design principles, compatibility. Automatic documentation and mathematical linguistics (New York), Vol. 12, No. 1, 1978, pp. 1-13. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 12, No. 1, 1978, pp. 1-10.]

77. CHILDS, G. H. L. The Prestel viewdata system. In: CSABA, L.; SZENTIVÁNYI, T.; TARNAY, K., eds. Networks from the user's point of view (proceedings of the IFIP TC-6 working conference COMNET '81, Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 603-615.

78. CHITNIS, A. M.; COSTA, J. M. Videotex services: network and terminal alternatives. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 485-497.

79. CHORAFAS, D. N. Data communications for distributed information systems. New York, Petrocelli Books, 1980, XII, 235p.

80. CICIORA, W. S. The role of the television receiver manufacturer in the United States. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 533-546.

81. CLARK, D.; POGRAN, K.; REED, D. An introduction to local area networks. IEEE proceedings (New York), Vol. 66, No. 11, 1978, pp. 1497-1517.

82. CLARK, G. E., JR. Computer and data communications standards organizations. In: Proceedings of computer networking symposium, Gaithersburg, MD, 15 December 1977. New York, Institute of Electrical and Electronics Engineers, 1977, pp. 1-8.

83. CLARKE, K. E.; CHILDS, G. H. L. The future of videotex. *Electronic publishing review* (Oxford), Vol. 1, No. 1, 1981, pp. 43-51.

84. CLEMENT, H.E.A. The automated authority file at the National Library of Canada. *International cataloguing* (London), Vol. 9, No. 4, 1980, pp. 45-48.

85. COATES, E. J. The Broad System of Ordering. *International forum on information and documentation* (Moscow), Vol. 4, No. 3, 1979, pp. 3-6.

86. COATES, E. J. The Broad System of Ordering. *International forum on information and documentation* (Moscow), Vol. 6, No. 1, 1981, pp. 24-30.

87. COATES, E. J. Switching languages for indexing. *Journal of documentation* (London), Vol. 26, No. 2, 1970, pp. 102-110.

88. COCHRANE, P. A. Can a standard for an online common command language be developed? *Online* (Weston, Conn.), Vol. 7, No. 1, 1983, pp. 36-37.

89. COHEN, D. On interconnection of computer networks. *In*: BEAUCHAMP, K. G., *ed*. *Interlinking of computer networks; proceedings of the NATO Advanced Study Institute held at Bonas, France, 28 August - 8 September 1978*. Dordrecht, D. Reidel, 1979, pp. 175-183.

90. COLLIER, M. Centralization and after: a review of the prospects for distributed processing in libraries. *Aslib proceedings* (London), Vol. 34, No. 6/7, 1982, pp. 295-300.

91. CONGER, L. D. *Online command chart*. Weston, Conn., Online, Inc., 1980, 9p.

92. COSTIGAN, D. M. *Electronic delivery of documents and graphics*. New York, Van Nostrand Reinhold, 1978, XVII, 344p.

93. COTTON, I. W. *Computer network interconnection: problems and prospects*. Washington, DC, Institute for Computer Sciences and Technology, National Bureau of Standards, 1977, VI, 73p. (NBS Special publication 500-6)

94. COUSINS, T. R.; DOMINICK, W. D. The management of data bases of data bases. *Proceedings of the American Society for Information Science* (White Plains, NY), Vol. 15, 1978, pp. 75-78.

95. CRAWFORD, W. CRT terminal checklist. *Journal of library automation* (Chicago), Vol. 13, No. 1, 1980, pp. 36-44.

96. CRAWFORD, W. EBCDIC bibliographic character sets--sources and uses: a brief report. *Journal of library automation* (Chicago), Vol. 12, No. 4, 1979, pp. 380-383.

97. CRINER, K.; JOHNSON-HALL, M. Videotex: threat or opportunity. *Special libraries* (New York), Vol. 71, No. 9, 1980, pp. 379-385.

98. CRONIN, B. Databanks. *Aslib proceedings* (London), Vol. 33, No. 6, 1981, pp. 243-250.

99. CRYSTAL, M. I.; JAKOBSON, G. E. FRED, a front end for databases. *Online* (Weston, Conn.), Vol. 6, No. 5, 1982, pp. 27-30.

100. DAHLBERG, I. The Broad System of Ordering (BSO) as a basis for an integrated social sciences thesaurus? *International classification* (Munich), Vol. 7, No. 2, 1980, pp. 66-72.

101. DAHLBERG, I. Conceptual definitions for INTERCONCEPT. *International classification* (Munich), Vol. 8, No. 1, 1981, pp. 16-22.

102. DAHLBERG, I. Major developments in classification. *Advances in librarianship* (New York), Vol. 7, 1977, pp. 41-103.

103. DAHLBERG, I. Toward establishment of compatibility between indexing languages. *International classification* (Munich), Vol. 8, No. 2, 1981, pp. 86-91.

104. DATATROL CORPORATION. *Common vocabulary approaches for government scientific and technical informations systems*. Silver Spring, Md., 1963.

105. DAVIS, C. H.; RUSH, J. E. *Guide to information science*. Westport, Conn., Greenwood Press, 1979, XVII, 305p.

106. DAVIS, N. C.; GOODMAN, S. E. The Soviet bloc's unified system of computers. *ACM computing surveys* (New York), Vol. 10, No. 2, 1978, pp. 93-122.

107. DAY, A. C. *Compatible Fortran*. Cambridge, Cambridge University Press, 1978, VII, 107p.

108. DAYTON, D. L.; LUNDEEN, J. W.; POLLACK, J. J. Automated techniques for online search guidance: a review. *In*: *4th international online information meeting, London, 9-11 December 1980*. Oxford, Learned Information, 1980, pp. 317-333.

109. DEBESSE, B. Multilingual terminology. *In*: *Overcoming the language barrier*: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 2. Munich, Verlag Dokumentation, 1977, pp. 129-134.

110. DEHNING, W.; ESSIG, H.; MAASS, S. *The adaptation of virtual man-computer interfaces to user requirements in dialogs*. Berlin, Springer-Verlag, 1981, X, 142p. (Lecture notes in computer science No. 110)

111. DERFALL, A. *Artificial intelligence as applied to input and output in the office or...making computers read and speak*. Cambridge, Mass., Kurzweil Computer Products, 1979, 8p.

112. DESJARDINS, R. ISO/ANSI reference model of open systems interconnection. *In*: *Trends & applications: computer network protocols*. New York, Institute of Electrical and Electronics Engineers, 1980, pp. 47-53.

113. *DEVSIS: the preliminary design of an International Information System for the Development Sciences*. Ottawa, International Development Research Centre, 1975, 247p.

114. DEXTRE, S. G.; CLARKE, T. M. A system for machine-aided thesaurus construction. *Aslib proceedings* (London), Vol. 33, No. 3, 1981, pp. 102-112.

115. DIENER, R. A. V.; TSUFFIS, M. L. A method for evaluating the feasibility of the intermediate lexicon concept for existing information retrieval systems. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 14, 1977, p. 28. (Full paper on microfiche)

116. DIERICKX, H. Personal communication. Paris, December 1982.

117. DIERICKX, H. A proposed common format for existing and projected computerized bibliographic information systems in the Caribbean. Paris, Unesco, 1982, IV, 78p. (FMR/PGI/83/159)

118. DIERICKX, H. The UNISIST reference manual for machine-readable bibliographic descriptions: current and anticipated developments. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 147-161.

119. DIERICKX, H.; HOPKINSON, A. Reference manual for machine-readable descriptions of research projects and institutions. Paris, Unesco, 1982, 174p. (PGI-82/WS/10)

120. DIERICKX, H.; HOPKINSON, A., eds. Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, 214p.

121. DIPPEL, G.; HOUSE, W. C. Information systems: data processing and evaluation. Glenview, IL, Scott, Foresman, 1969, XI, 493p.

122. DODD, S. A. Bibliographic references for numeric social science data files: suggested guidelines. Journal of the American Society for Information Science (New York), Vol. 30, No. 2, 1979, pp. 77-82.

123. DODD, S. A. Toward integration of catalog records on social science machine-readable data files into existing bibliographic utilities: a commentary. Library trends (Champaign, IL), Vol. 30, No. 3, 1982, pp. 335-361.

124. DOSZKOCS, T. E. AID, an associative interactive dictionary for online searching. Online review (Oxford), Vol. 2, No. 2, 1978, pp. 163-173.

125. DOSZKOCS, T. E.; RAPP, B. A. Searching MEDLINE in English: a prototype user interface with natural language query, ranked output, and relevance feedback. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 16, 1979, pp. 131-139.

126. DOSZKOCS, T. E.; RAPP, B. A.; SCHOOLMAN, H. M. Automated information retrieval in science and technology. Science (Washington, DC), Vol. 208, No. 4439, 1980, pp. 25-30.

127. DUCHESNE, R. M. MARC and SUPERMARC. In: The exchange of bibliographic data and the MARC format. Munich, Verlag Dokumentation, 1972, pp. 37-57.

128. DUCHESNE, R. M.; ISLAM, M.; GUENTER, D. A. Overview of computerized library networking in Canada. Ottawa, National Library of Canada, 1979, IX, 86p.

129. DUCROT, J. M. The TITUS II system. NEWSIDIC, April 1974, pp. 4-6.

130. DURANCE, C. J.; BUCHINSKI, E. J.; GUENTER, D. A. iNet and Canadian libraries: new telecommunications facilities for library and information services. *Canadian journal of information science* (Toronto), Vol. 7, No. 1, 1982, pp. 1-10.

131. DUSOULIER, N. Compatibilité et normalisation. *Information et documentation* (Paris), Vol. 4, December, 1975, pp. 41-46.

132. EADIE, D. *A user's guide to computer peripherals*. Englewood Cliffs, NJ, Prentice-Hall, 1982, X, 244p.

133. EAST, H.; MARTINELLI, M. T. Building on what exists: the development of AGRIS as an example of systems interconnection. In: *Information systems: their interconnection and compatibility*. Vienna, International Atomic Energy Agency, 1975, pp. 41-50.

134. EASTLAKE, D. E.; LOZANO-PEREZ, T.; LOW, D. A. *Design of the version I prototype chemical substances information network*. Cambridge, Mass., Computer Corporation of America, 1980, VII, 231p. (Technical Report No. CCA-80-6)

135. EDELHART, M. Optical discs: the omnibus medium. *Technology* (Boulder, CO), Vol. 1, No. 1, 1981, pp. 42-57.

136. EDSTRÖM, M. E.; WALLIN, M. A. Implementation of a new interactive on-line information retrieval system: 3RIP. In: *1st international online information meeting, London, 13-15 December 1977*. Oxford, Learned Information, 1977, pp. 83-93.

137. EHLERS, H. J. Legibility aspects of coded online information. In: *5th international online information meeting, London, 8-10 December 1981*. Oxford, Learned Information, 1981, pp. 455-467.

138. EHLERS, H. J. *Numbering systems of ISO/TC 46: status of work*. Paper presented to the ISBN Advisory Panel, Berlin, October 1981, 14p.

139. ELIEZER, I. Use of computers in handling of laboratory data. In: ROSSMASSLER, S. A.; WATSON, D. G., eds. *Data handling for science and technology: an overview and sourcebook*. Amsterdam, North-Holland, 1980, pp. 115-129.

140. ELLMANN, D. Stav a vysledky prace s tezaury v soustave VTI NDR. *Ceskoslovenska informatika* (Prague), Vol. 18, No. 7/8, 1976, pp. 193-195.

141. EMMONS, W. F. Data network protocol standards. In: *Proceedings of computer networking symposium, Gaithersburg, MD, 15 December 1977*. New York, Institute of Electrical and Electronics Engineers, 1977, pp. 9-17.

142. ENGEL, S. E.; GRANDA, R. E. *Guidelines for man/display interfaces*. Poughkeepsie, NY, IBM Poughkeepsie Laboratory, 1975, 42p. (TR 00.2720)

143. ENGER, N. L. *Documentation standards for computer systems*. Fairfax Station, VA, Technology Press, 1976, VIII, 231p.

144. EPSTEIN, H. The technology of library and information networks. *Journal of the American Society for Information Science* (New York), Vol. 31, No. 6, 1980, pp. 425-437.

145. Études préalables a l'élaboration d'un code universel pour la présentation et le traitement de textes multilingues dans un contexte européén. Aix-en-Provence, CREL-France, 1981, 27p.

146. Euronet - DIANE user interface box technical specification. 11p.

147. EUROPEAN COMMUNITIES. Report drawn up on behalf of the Committee on Economic and Monetary Affairs on the proposals from the Commission of the European Communities to the Council (Doc. 1-431/81). European Parliament working documents 1982-1983. 1982, 13p. (Document 1-41/82)

148. EXPERT GROUP TO ADVISE ACC ON THE CO-ORDINATION OF INFORMATION SYSTEMS IN THE UNITED NATIONS FAMILY. Final report on strengthening the co-ordination of information systems in the United Nations family. New York, United Nations, May 1982, 35p.

149. FANGMEYER, H. Semi-automatic indexing: state of the art. Neuilly sur Seine, NATO, Advisory Group for Research and Development, 1974.

150. FAYEN, E. G.; COCHRAN, M. A new user interface for the Dartmouth on-line catalog. In: WILLIAMS, M E.; HOGAN, T. H. comps. National online meeting proceedings, New York, 31 March - 1 April 1982. Medford, NJ: Learned Information, 1982, pp. 87-97.

151. FELBER, H. International efforts to overcome difficulties in technical communication. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 86-99.

152. FELBER, H. A plan for data management within Termnet. In: Terminological data banks. Munich, K. G. Saur, 1980, pp. 190-196. (Infoterm series 5)

153. FELBER, H. A study of terminological data elements on the basis of the general theory of terminology and a comparative study of those data elements used in terminological data banks all over the world. Vienna, Infoterm, 1981, 40p. (Draft)

154. FELBER, H.; GALINSKI, CH.; NEDOFITY, W. Ten years of Infoterm, a report on activities and achievements. In: Terminologies for the eighties. Munich, K. G. Saur, 1982, pp. 19-95. (Infoterm series 7)

155. FLORES, I. Codes. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 210-215.

156. FOLTS, H. C. Interface standards for public data networks. In: Proceedings of computer networking symposium, Gaithersburg, MD, 15 December 1977. New York, Institute of Electrical and Electronics Engineers, 1977, pp. 18-25.

157. FOLTS, H. C.; KARP, H. R., eds. McGraw-Hill's compilation of data communications standards. New York, McGraw-Hill, 1978, 1133p.

158. FOSDICK, H. Computer basics for librarians and information scientists. Arlington, Information Resources Press, 1981, X, 203p.

159. FOSDICK, H. Distributed data processing: what is it, and what does it mean. Illinois libraries (Springfield, IL), Vol. 63, No. 8, 1981, pp. 562-572.

160. FOSDICK, H. The microcomputer catalyst. In: SMITH, L. C., ed. New information technologies--new opportunities (papers presented at the 1981 Clinic on Library Applications of Data Processing, 26-29 April 1981). Urbana-Champaign, University of Illinois, Graduate School of Library and Information Science, 1982, pp. 13-27.

161. FOSKETT, A. C. The Broad System of Ordering: old wine into new bottles? International forum on information and documentation (Moscow), Vol. 4, No. 3, 1979, pp. 7-12.

162. FOX, M. S.; PALAY, A. J. Machine-assisted browsing for the naive user. In: DIVILBISS, J. L., ed. Public access to library automation (papers presented at the 1980 Clinic on Library Applications of Data Processing, 20-23 April 1980). Urbana-Champaign, University of Illinois, Graduate School of Library and Information Science, 1981, pp. 77-98.

163. FRANK, H.; HILGERS, R.; MEYER, I. Documentation in a constructed language and man-machine communication. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 33-37.

164. FRANZMEIER, G. Authority file for corporate authors in the German Serials Data Base. In: HELAL, A. H.; WEISS, J. W., eds. Current trends in serials automation. Essen, Essen University Library, 1981, 201p.

165. FRENCH, E. J. Standardization as a factor in information transfer. Journal of information science (Amsterdam), Vol. 3, No. 2, 1981, pp. 91-100.

166. FRIED, J. B. On-line numeric data bases. Bulletin of the American Society for Information Science (Washingtin, DC), Vol. 1, No. 7, 1975, pp. 17-18.

167. GAAZE-RAPOPORT, M. G. A unified system of information languages: dreams and some problems. Automatic documentation and mathematical linguistics (New York), Vol. 12, No. 1, 1978, pp. 45-48. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 12, No. 1, 1978, pp. 32-34.]

168. GABLER, H. G. Advanced text communication with teletex. In: CSABA, L.; SZENTIVÁNYI, T.; TARNAY, K., eds. Networks from the user's point of view (proceedings of the IFIP TC-6 working conference COMNET '81. Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 591-601.

169. GARDIN, N. Intermediate lexicon: a new step towards international cooperation in scientific and technical information. UNESCO bulletin for libraries (Paris), Vol. 23, No. 2, 1969, pp. 58-63.

170. GARFIELD, E. Incomplete citations and other sources of bibliographic chaos. In: GARFIELD, E., ed. Essays of an information scientist. Vol. 1. Philadephia, ISI Press, 1977, p. 40.

171. GAROOGIAN, R. Pre-written software: identification, evaluation and selection. Software review (Westport, Conn.), Vol. 1, No. 1, 1982, pp. 11-34.

172. GEBHARDT, F.; STELLMACHER, I. Opinion paper: design criteria for documentation retrieval languages. Journal of the American Society for Information Science (New York), Vol. 29, No. 4, 1978, pp. 191-199.

173. GEE, K. C. E. An introduction to open systems interconnection. Manchester, National Computing Centre, 1980, 61p.

174. GELLER, V. J.; LESK, M. E. User interfaces to information systems: choices vs. commands. ACM SIGIR forum (New York), Vol. 17, No. 4, 1983, pp. 130-135.

175. GIBB, J. M.; PHILLIPS, E. Scientific and technical publishing in a multilingual society. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 13-27.

176. GIERTZ, L. M. The SfB system: a case study on an information system for global co-operation in the field of building construction. In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 349-355.

177. GIULIANO, V. E. The mechanization of office work. Scientific American (New York), Vol. 247, No. 3, 1982, pp. 148-164.

178. GLUSHKOV, V. M.; SKOROKHOD'KO, E. F.; STROGNII, A. A. Evaluation of the degree of compatibility of information retrieval languages of document retrieval systems. Automatic documentation and mathematical linguistics (New York), Vol. 12, No. 1, 1978, pp. 18-26. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 12, No. 1, 1978, pp. 14-19.]

179. GOETSCHALCKX, J. Automatic translation. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 2. Munich, Verlag Dokumentation, 1977, pp. 142-146.

180. GOETSCHALCKX, J. Terminological activities in the European institutions with special reference to Eurodicautom. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 123-152.

181. GOLDSTEIN, C. M. Optical disk technology and information. Science (Washington, DC), Vol. 215, No. 4534, 1982, pp. 862-868.

182. GOLDSTEIN, C. M.; FORD, W. H. The user-cordial interface. Online review (Oxford), Vol. 2, No. 3, 1978, pp. 269-275.

183. GOPINATH, M. A.; PRASAD, K. N. Compatibility of the principles for design of thesaurus and classification scheme. Library science with a slant to documentation (Bangalore), Vol. 13, No. 2, 1976, pp. 56-66.

184. GORMAN, M. Bibliographic standardization and machine records. In: COWARD, R. E.; YELLAND, M., eds. The interchange of bibliographic information in machine readable form. London, Library Association, 1975, pp. 86-92.

185. GOSIN, I. YA.; SMIRNOVA, M. N. Novyi etap vrazvitii teorii i praktiki bibliograficheskogo opisaniya proizvedenii pechati v SSSR. Nauchnye i tekhnicheskie biblioteki SSSR (Moscow), Vol. 4, 1978, pp. 3-10.

186. GOVE, N. B.; HUGHES, T. E.; MATIUSHIN, T. D.; TURTANOV, N. V. A common communication format for bibliographic data exchange. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 89-102.

187. GOVE, N. et al Use of ISO standards in INIS. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 335-347.

188. GRAY, J. C. Acceptance of international standards. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 301-305.

189. GRIFFITH, W. G. Advances in merging full text and videotex features for information retrieval. In: WILLIAMS, M. E.; HOGAN, T. H., comps. National online meeting proceedings, New York, 31 March - 1 April 1982. Medford, N.J., Learned Information, 1982, pp. 159-161.

190. GROENEWEGEN, H. W.; MARSHALL, J. Optical character recognition: use of OCR techniques in decentralized data collection for bibliographic information systems. Ottawa, International Development Research Centre, 1976, 96p. (IDRC-TS3)

191. Guidelines for authority and reference entries. Draft prepared by the IFLA Working Group on an International Authority System, October 1982.

192. Guidelines for the national bibliographic agency and the national bibliography. Prepared by the IFLA International Office for UBC. Paris, Unesco, 1979, 72p. (PGI/79/WS/18)

193. GUTIERREZ, F.; REY, A. Critical approach to the output formats of on-line available data bases. In: 1st international on-line information meeting, London, 13-15 December 1977. Oxford, Learned Information, 1977, pp. 103-113.

194. HABER, R. Euronet and videotex: competitive or complementary? Computer communications (Guildford), Vol. 4, No. 3, 1981, pp. 99-105.

195. HAMAN, P. Formy záznamu v informačnich pramenech a jejich konvertibilita. Ceskoslovenska informatika (Prague), Vol. 18, No. 4, 1976, pp. 103-109.

196. HAMMOND, W. Satellite thesaurus construction. In: The thesaurus in action: background information for a thesaurus workshop at the 32nd annual convention of the American Society for Information Science. San Francisco, American Society for Information Science, 1969, pp. 14-20. (AD 694 590)

197. HAMMOND, W. Statistical association methods for simultaneous searching of multiple document collections. In: STEVENS, M. E.; GIULIANO, V. E.; HEILPRIN, L. B., eds. Statistical association methods for mechanized documentation: symposium proceedings. Washington, DC, National Bureau of Standards, 1965, pp. 237-243. (NBS Miscellaneous Publication 269)

198. HAMMOND, W.; ROSENBORG, S. Experimental study of convertibility between large technical indexing vocabularies, with a table of indexing equivalents. Silver Spring, Maryland, 1962.

199. HAMMOND, W.; ROSENBORG, S. Indexing terms of announcement publications for government scientific and technical research reports: a composite vocabulary. Silver Spring, Maryland, Datatrol Corporation, 1964. 2 volumes.

200. HARPER, L. G. A comparative review of BRS, DIALOG and ORBIT. Reference services review (Ann Arbor), Vol. 9, No. 1, 1981, pp. 39-51.

201. HAWKINS, D. T. Machine-readable output from online searches. Journal of the American Society for Information Science (New York), Vol. 32, No. 4, 1981, pp. 253-256.

202. HEATH, T. Alternative videodisc systems. Videodisc/videotex (Westport, Conn.), Vol. 1, No. 4, 1981, pp. 228-237.

203. HECHT, J. Fiber optics for short-distance transmission. Technology (Boulder, CO), Vol. 2, No. 2, 1982, pp. 40-52.

204. HECHT, J. Lasers store a wealth of data. High technology (Boston), Vol. 2, No. 3, 1982, pp. 60-67.

205. HEILIK, J.; BURTON, R. H. CASSI and the compression of journal names in an information retrieval system. Journal of chemical information and computer sciences (Washington, DC), Vol. 17, No. 4, 1977, pp. 210-211.

206. HELLER, S. R.; MILNE, G. W. A. Linking scientific databases--the NIH-EPA chemical information system. Online (Weston, Conn.), Vol. 5, No. 4, 1980, pp. 45-56.

207. HELLER, S. R.; MILNE, G. W. A. The NIH-EPA Chemical Information System (CIS). In: GLAESER, P. S., ed. Data for science and technology: Proceedings of the Seventh International CODATA Conference, Kyoto, Japan, 8-11 October 1980. Oxford, Pergamon Press, 1981, pp. 343-346.

208. HELLERMAN, H.; SMITH, I. A. Computer systems. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 318-323.

209. HENDERSON, M. M. Standards: developments and impacts. Special libraries (New York), Vol. 72, No. 2, 1981, pp. 142-148.

210. HENDERSON, M. M. Study of cooperation in the processing of the technical report literature. Unpublished report prepared for the National Technical Information Service, 1980, 44p.

211. HENDERSON, M. M.; MOATS, J. S.; STEVENS, M. E.; NEWMAN, S. M. Cooperation, convertibility, and compatibility among information systems: a literature review. Washington, DC, National Bureau of Standards, 1966, III, 140p. (NBS Miscellaneous Publication 276)

212. HENDRIX, G. G.; SACERDOTI, E. D.; SAGALOWICZ, D.; SLOCUM, J. Developing a natural language interface to complex data. ACM transactions on database systems (New York), Vol. 3, No. 2, 1978, pp. 105-147.

213. HENSHAW, M. Conversion sampler. Library journal (New York), Vol. 92, No. 19, 1967, pp. 3964-3966.

214. HESSELAGER, L. En introduktion til ISDS. Nordisk tidskrift foer bok- och biblioteksvaesen (Stockholm), Vol. 65, No. 1, 1978, pp. 1-17.

215. HEYMANS, F. How human-usable is interchangeable? Or, shall we produce catalogues or babelographic towers? Library resources and technical services (Chicago), Vol. 26, No. 2, 1982, pp. 156-169.

216. HEYS, F. Private videotex systems and access via public services. In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 199-210.

217. HICKEY, D. J. Bibliographic control in theory. IFLA journal (Munich), Vol. 6, No. 3, 1980, pp. 234-241.

218. HICKEY, T. B.; SPIES, P. B. Standards for information display. Library trends (Champaign, IL), Vol. 31, No. 2, 1982, pp. 315-324.

219. HILDRETH, C. R. Online public access catalogs: the user interface. Dublin, OH, OCLC, 1982, XV, 263p.

220. HILL, J. S. The Northwestern Africana project: an experiment in decentralized bibliographic and authority control. College and research libraries (Chicago), Vol. 42, No. 4, 1981, pp. 326-332.

221. HITZEROTH, C. MADOK--the German magnetic tape exchange format in the field of documentation. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 73-75.

222. HÖPKER, W. W. Struktur und Kompatibilitat des Thesaurus der Medizin. International classification (Munich), Vol. 3, No. 2, 1976, pp. 81-84.

223. HOPKINSON, A. Personal communication. 1983.

224. HOPKINSON, A. The use made of the UNISIST reference manual: a brief history. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 49-51.

225. HORNE, C. Gateway--an enhancement of Prestel. Aslib proceedings (London), Vol. 34, No. 5, 1982, pp. 266-270.

226. HOROWITZ, A. J.; BERGMAN, R. F. Improved productivity in multi-system information retrieval through the CSIN intelligent terminal. In: WILLIAMS, M. E.; HOGAN, T. H., comps. National online meeting proceedings. Medford, N.J., Learned Information, 1981, pp. 287-292.

227. HOROWITZ, A. J.; EASTLAKE, D. E. III; LOW, D. An intelligent terminal for access to heterogeneous chemical information systems. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 17, 1980, pp. 83-85.

228. HORSNELL, V. The intermediate lexicon: an aid to international co-operation. Aslib proceedings (London), Vol. 27, No. 2, 1975, pp. 57-66.

229. HORSNELL, V. Intermediate lexicon for information science: a feasibility study. London, Polytechnic of North London, School of Librarianship, 1974, 110p.

230. HORSNELL, V.; MERRETT, A. Intermediate lexicon research project. Phase 2. London, Polytechnic of North London, School of Librarianship, 1978, 100p.

231. HOYT, D. R. Word processor applications at the USDA's Technical Information Systems. In: SMITH, L. C., ed. New information technologies--new opportunities (papers presented at the 1981 Clinic on Library Applications of Data Processing, 26-29 April 1981). Urbana-Champaign, University of Illinois, Graduate School of Library and Information Science, 1982, pp. 43-59.

232. HULME, E. W. Statistical bibliography in relation to the growth of modern civilization. London, Grafton, 1923, 44p.

233. HUTH, E. J. New forms for references. British medical journal (London), No. 6179, 1979, pp. 1697-1698.

234. ICSU AB/CODATA JOINT WORKING GROUP ON TAGGING AND FLAGGING. Flagging and tagging data to indicate its presence and facilitate its retrieval. Information reports and bibliographies (New York), Vol. 6, No. 4, 1977, pp. 7-25.

235. ILJON, A. Creation of thesauri for EURONET. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 417-437.

236. ILJON, A. Scientific and technical data bases in a multilingual society. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 28-32.

237. Inclusion of nonroman character sets. Journal of library automation (Chicago), Vol. 14, No. 3, 1981, pp. 210-215.

238. iNet newsletter (Ottawa), Vol. 1, No. 1, 1982.

239. INFORMATION SCIENCES OF AUSTRALIA. Local interfaces to the Australian bibliographic network. Canberra, National Library of Australia and Australian Advisory Council on Bibliographical Services, 1982, II, 127p.

240. INTERNATIONAL INSTITUTE OF COMMUNICATIONS. The use of satellite communication for information transfer. Paris, Unesco, 1982, IV, 125p. (PGI-82/WS/5)

241. International list of periodical title word abbreviations. Paris, International Council of Scientific Unions Abstracting Board, 1970, XII, 39p.

242. The international MARC network: bibliographic study. London, International Federation of Library Associations and Institutions, 1977, 28p.

243. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Transliteration of Slavic Cyrillic characters. Geneva, ISO, 1968. (ISO/9)

244. ISBD(G): general international standard bibliographic description. London, IFLA International Office for UBC, 1977, X, 24p.

245. ISBD(M): international standard bibliographic description for monographic publications. London, IFLA International Office for UBC, 1978, VIII, 51p.

246. ISBD(S): international standard bibliographic description for serials. London, IFLA International Office for UBC, 1977, VIII, 61p.

247. ISO standards handbook. 1. Information transfer. Second edition. Geneva, International Organization for Standardization, 1982.

248. It's SUPERINDEX. CHRONOLOG (Palo Alto), Vol. 10, No. 5, 1982, pp. 85-86.

249. JACHOWICZ, R. L. Application of classification as a basis for the formulation of a thesaurus. In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 356-362.

250. JACOBSEN, T. Network independent, distributed systems. In: CSABA, L.; SZENTIVÁNYI, T.; TARNAY, K., eds. Networks from the user's point of view (proceedings of the IFIP TC-6 working conference COMNET '81, Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 305-314.

251. JACOBUS, D. P.; ZABRISKIE, K. H.; GORDON, M. Compatibility in chemical information systems. Journal of chemical documentation (Washington, DC), Vol. 9, No. 2, 1969, pp. 118-125.

252. JAMIESON, S. H.; ODDY, R. N. Low-cost implementation of experimental retrieval techniques for online users. In: 4th international online information meeting, London, 9-11 December 1980. Oxford, Learned Information, 1980, pp. 201-209.

253. JASTER, J. Subsection scheme for dictionary of indexing equivalents. Silver Spring, Md., Datatrol Corporation, 1963.

254. JEDLIČKA, B.; STARÝ, M.; VERNER, V. Jednotný automatizovaný systém VTEI pro československé plynarenství. Ceskoslovenska informatika (Prague), Vol. 18, No. 2, 1976, pp. 39-47.

255. JENNINGS, P. G.; NEWMAN, L. E.; WILKINSON, W. B. Data capture by optical scanning of published material for database enhancement. Program (London), Vol. 16, No. 1, 1982, pp. 17-26

256. JONES, C. L. Linking bibliographic data bases: a discussion of the Battelle technical report. Washington, DC, Council on Library Resources, 1980, 28p.

257. JONES, K. P. Problems associated with the use of compound words in thesauri with special reference to BS 5723: 1979. Journal of documentation (London), Vol. 37, No. 2, 1981, pp. 53-68.

258. KARSHMER, A. I. Front-end processors. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 580-582.

259. KELLY, P. T. F. The EURONET-DIANE project. International forum on information and documentation (Moscow), Vol. 7, No. 2, 1982, pp. 22-27.

260. KENNEY, B. L. Basics of telecommunications. In: Telecommunications and libraries: a primer for librarians and information managers. White Plains. NY, Knowledge Industry Publications, 1981, pp. 23-40.

261. KENT, A. K. Problems of remote processing of magnetic tape information files. Aslib proceedings (London), Vol. 20, No. 11, 1968, pp. 502-508.

262. KESNER, R. M.; HURST, D. MARS: a development report. ADPA (Liege), Vol. 3, No. 2, 1980, pp. 11-20.

263. KIMBLETON, S. R.; SCHNEIDER, G. M. Computer communications networks: approaches, objectives, and performance considerations. ACM computing surveys (New York), Vol. 7, No. 3, 1975, pp. 129-173.

264. KING, M.; PERSCHKE, S. EUROTRA and its objectives. Multilingua (Amsterdam), Vol. 1, No. 1, 1982, pp. 27-32.

265. KINNUCAN, P. Local networks battle for billion-dollar market. High technology (Boston), Vol. 1, No. 2, 1981, pp. 64-72.

266. KIRSTEIN, P. T. Some international developments in data services. In: BEAUCHAMP, K. G., ed. Interlinking of computer networks; proceedings of the NATO Advanced Study Institute held at Bonas, France, 28 August - 8 September 1978. Dordrecht, D. Reidel, 1979, pp. 3-30.

267. KIRTLAND, M.; ATHERTON, P.; HARPER, S. Integrating borrowed records into a data base: impact on thesaurus development and retrieval. Database (Weston, Conn.), Vol. 3, No. 4, December 1980, pp. 26-33.

268. KODOLA, V. E.; SKRIPKIN, V. A. Compatibility problems in the CMEA countries' international scientific and technical information system. International forum on information and documentation (Moscow), Vol. 3, No. 2. 1978, pp. 18-23.

269. KOEL, Å. I. Bibliographic control at the crossroads: do we get our money's worth? Journal of academic librarianship (Boulder, CO), Vol. 7, No. 4, 1981, pp. 220-222.

270. KOHL, E. Areas of possible incompatibilities in the international exchange of bibliographic information in machine-readable form based on a compatibility study for MAB 1 data. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April 1978. Budapest, OMKDK-Technoinform, 1978, pp. 52-58.

271. KOHL, E. Character sets and coding techniques. In: COWARD, R. E.; YELLAND, M., eds. The interchange of bibliographic information in machine readable form. London, Library Association, 1975, pp. 93-94. (Library Association research publication No. 17)

272. KRÍŽIKOVÁ, S. Medzinárodný systém vedeckých a technických informácií pre vodňe hospodarstvo - vodoinform. Ceskoslovenska informatika (Prague), Vol. 18, No. 11, 1976, pp. 302-304.

273. KRUMMEL, D. W.; HOWELL, J. B. Bibliographic standards and style. Scholarly publishing (Toronto), Vol. 10, No. 3, 1979, pp. 223-240.

274. KUIPERS, J. W.; THORPE, R. W. Information management problems relating to numeric data base services for DOE. Bedford, Mass., QEI Inc., 1979, VIII, 46p.

275. LABADI, A. IIASA gateway system and experiments in daily operation. In: CSABA, L.; SZENTIVÁNYI, T.; TARNAY, K., eds. Networks from the user's point of view (proceedings of the IFIP TC-6 working conference COMNET '81, Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 127-131.

276. LAGUEUX, P. B. Standards for networks and the identification of some missing links. In: MARKUSON, B. E.; WOOLLS, B., eds. Networks for networkers: critical issues in cooperative library development. New York, Neal-Schuman, 1980, pp. 174-184.

277. LAKE, A. CSIN user support: intelligent terminal study. Arlington, VA, Bolt Beranek and Newman, 1982, IX, 88p. (BBN report no. 4838)

278. LARSON, J.; BERNHARD, G. A.; PADIN, M. E. Comparison of printed bibliographic descriptions distributed by BIOSIS, CAS, and Ei. Journal of the American Society for Information Science (Washington, DC), Vol. 27, No. 1, 1976, pp. 46-52.

279. LARSSON, R. On implementing the common command language within 3RIP. Stockholm, Royal Institute of Technology Library, 1980, 36p. (Report TRITA-LIB-4067)

280. LEBOWITZ, A. A common communication format for machine-readable bibliographic data. In: DIERICKX, H.; HOPKINSON, A., eds. Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April 1978. Budapest, OMKDK-Technoinform, 1978, pp. 195-208.

281. LEE, J. A.; RALSTON, A. Character set. In: RALSTON, A.; MEEK, C. L.; eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 206-208.

282. LERNER, R. G. Access to data in the primary literature. In: ROSSMASSLER, S. A.; WATSON, D. G., eds. Data handling for science and technology: an overview and sourcebook. Amsterdam, North-Holland, 1980, pp. 75-82.

283. LE ROUX, Y. Report of the first meeting of ISO/TC97/SC18--text preparation and interchange. Computers & standards (Amsterdam), Vol. 1, No. 1, 1982, pp. 23-25.

284. LESK, M. Combining data bases: national and cartographic files. Murray Hill, NJ, Bell Laboratories, 1981, 9p.

285. LEVINE, G. R. Alternative database structures for bibliographic information retrieval systems--their past and future. In: WILLIAMS, M. E.; HOGAN, T. H., comps. National online meeting proceedings, New York, 24-26 March 1981. Medford, NJ, Learned Information, 1981, pp. 331-339.

286. LEVINE, G. R. Developing databases for online information retrieval. Online review (Oxford), Vol. 5, No. 2, 1981, pp. 109-120.

287. LEVY, F. Compatibility between classifications and thesauri: evaluation of a first study in the field of information storage and retrieval. Paper presented at the International Federation for Documentation conference, Tokyo, 1967.

288. LITTLE, J. L. A computer network protocol at the application level for libraries and other information services. Journal of library automation (Chicago), Vol. 11, No. 3, 1978, pp. 239-245.

289. LITVINTSEVA, L. V.; SENIN, G. V. A dialog language processor for interface with a data base model. Automatic documentation and mathematical linguistics (New York), Vol. 11, No. 1, 1977, pp. 64-74. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 11, No. 2, 1977, pp. 22-29.]

290. LLOYD, G. A. The UDC in its international aspects. Aslib proceedings (London), Vol. 21, No. 5, 1969, pp. 204-208.

291. LLOYD, G. A. The Universal Decimal Classification as an international switching language. In: WELLISCH, H.; WILSON, T. D., eds. Subject retrieval in the seventies: new directions. Westport, Conn., Greenwood Publishing Co., 1972, pp. 116-125.

292. LONG, P. L. Study of message text formats: bibliographic search queries. Washington, DC, Library of Congress, 1979, V, 28p. (Network planning paper No. 5)

293. LOWENTHAL, E. Database systems for local nets. Datamation (New York), Vol. 28, No. 9, 1982, pp. 97-106.

294. LU, C. Microcomputers: the second wave. High technology (Boston), Vol. 2, No. 5, 1982, pp. 36-52.

295. LUEDKE, J. A.; GABOR, J. K.; FRIED, J. B. Numeric data bases and systems. Annual review of information science and technology (White Plains, NY), Vol. 12, 1977, pp. 119-181.

296. MACKENZIE OWEN, J. S. Format incompatibility and the exchange of bibliographic information: a comparative study. Paris, Unesco, 1976, IV, 60p. (SC.76/WS/1)

297. MACKENZIE OWEN, J. Some fundamentals behind information exchange: a view on optimal solutions. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April 1978. Budapest, OMKDK-Technoinform, 1978, pp. 9-12.

298. MAGUIRE, M. An evaluation of published recommendations on the design of man-computer dialogues. International journal of man-machine studies (London), Vol. 16, No. 3, 1982, pp. 237-261.

299. MAHON, B. EURONET DIANE--the European on-line information network. Program (London), Vol. 14, No. 2, 1980, pp. 69-75.

300. MALCOM, J. B.; PROMISEL, J. L. A glossary for library networking. Washington, DC, Library of Congress, 1978, 34p. (Network planning paper No. 2)

301. MALINCONICO, S. M. Authority control in a bibliographic network environment. In: FURUYA, N. Y., ed. What's in a name? Control of catalogue records through automated authority files. Toronto, University of Toronto Library Automation System, 1978, pp. 161-178.

302. MALINCONICO, S. M. Bibliographic data base organization and authority file control. Wilson library bulletin (Bronx, NY), Vol. 54, No. 1, 1979, pp. 36-45.

303. MALINCONICO, S. M. The coordination of bibliographic control. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April 1978. Budapest, OMKDK-Technoinform, 1978, pp. 61-67.

304. MALINCONICO, S. M.; GRUTCHFIELD, W. R.; STEINER, E. J. Vernacular scripts in the NYPL automated bibliographic control system. Journal of library automation (Chicago), Vol. 10, No. 3, 1977, pp. 205-224.

305. MAMRAK, S. A. The design and development of resource-sharing services in computer communications networks: a survey. Advances in computers (New York), Vol. 16, 1977, pp. 183-219.

306. A manual of style for authors, editors and copywriters. Twelfth edition. Chicago, University of Chicago Press, 1969, 546p.

307. MARCUS, R. S. An automated expert assistant for information retrieval. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 18, 1981, pp. 270-273.

308. MARCUS, R. S. Search aids in a retrieval network. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 17, 1980, pp. 394-396.

309. MARCUS, R. S.; REINTJES, J. F. A translating computer interface for end-user operation of heterogeneous retrieval systems. I. Design. Journal of the American Society for Information Science (New York), Vol. 32, No. 4, 1981, pp. 287-303.

310. MARCUS, R. S.; REINTJES, J. F. A translating computer interface for end-user operation of heterogeneous retrieval systems. II. Evaluations. Journal of the American Society for Information Science (New York), Vol. 32, No. 4, 1981, pp. 304-317.

311. MARSH, F. E., JR. Videodisc technology. Journal of the American Society for Information Science (New York), Vol. 33, No. 4, 1982, pp. 237-244.

312. MARTI, B. Broadcast text information in France. In: Viewdata and videotext. 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 359-370.

313. MARTI, B.; SCHWARTZ, C. Videotex standardization: a French viewpoint. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 247-251.

314. MARTIN, T. H. A feature analysis of interactive retrieval systems. Stanford, Institute for Communication Research, Stanford University, 1974, III, 100p. (PB-235 952)

315. MATTHEWS, J. R. Online public access catalogs: assessing the potential. Library journal (New York), Vol. 107, No. 11, 1982, pp. 1067-1071.

316. MATTHEWS, J. R. 20 Qs & As on automated integrated library systems. American libraries (Chicago), Vol. 13, No. 6, 1982, pp. 367-371.

317. MCCRUM, W. A. Current CCITT standards. In: BEAUCHAMP, K. G., ed. Interlinking of computer networks; proceedings of the NATO Advanced Study Institute held at Bonas, France, 28 August-8 September 1978. Dordrecht, D. Reidel, 1979, pp. 95-110.

318. MCGLYNN, D. R. Distributed processing and data communications. New York, John Wiley & Sons, 1978, XIII, 305p.

319. MCGUIRE, J. S. The impact of PCM's on distributed processing. Datamation (New York), Vol. 25, No. 2, 1979, pp. 134-135, 138.

320. MCKEAN, J. M. Facsimile and libraries. In: Telecommunications and libraries: a primer for librarians and information managers. White Plains, NY, Knowledge Industry Publications, 1981, pp. 91-120.

321. MCMURDO, G. The interface between computerized retrieval systems and micrographic retrieval systems. Journal of information science (Amsterdam), Vol. 1, No. 6, 1980, pp. 345-349.

322. MEADOW, C. T. Matching users and user languages in information retrieval. Online review (Oxford), Vol. 5, No. 4, 1981, pp. 313-322.

323. MEADOW, C. T.; HEWETT, T. T.; AVERSA, E. S. A computer intermediary for interactive database searching. I. Design. Journal of the American Society for Information Science (New York), Vol. 33, No. 5, 1982, pp. 325-332.

324. MEEK, C. L. Compatibility. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 236-237.

325. MEHLMANN, M. When people use computers: an approach to developing an interface. Englewood Cliffs, NJ, Prentice-Hall, 1981, XVII, 142p.

326. MEIJER, A.; PEETERS, P. Computer network architectures. Rockville, Maryland, Computer Science Press, 1982, XI, 396p.

327. MELLEY, M. L. Environmental information and a data dictionary. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 17, 1980, pp. 351-353.

328. MERALI, Z. Word processing-output microform: after COM, can word processor/micro-/mini-computer-output microform take off? Journal of micrographics (Silver Spring, MD), Vol. 15, No. 6, 1982, pp. 18-24.

329. MERRILL, D.; AUSTIN, D. M., eds. ERDA Interlaboratory Working Group for Data Exchange (IWGDE) annual report for fiscal year 1976. Berkeley, University of California, Lawrence Berkeley Laboratory, 1976, 123p. (LBL-5329)

330. MEYROWITZ, N.; VAN DAM, A. Interactive editing systems: part II. ACM computing surveys (New York), Vol. 14, No. 3, 1982, pp. 353-415.

331. MICHAELS, A. S.; MITTMAN, B.; CARLSON, C. R. A comparison of the relational and CODASYL approaches to data-base management. ACM computing surveys (New York), Vol. 8, No. 1, 1976, pp. 125-151.

332. MICHEL, J. Personal communication. Paris, May 1982.

333. MIDDLETON, R. W.; EGGERT, J.; FRENCH, E. J. Application of standards in systems interconnection. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 307-333.

334. MILLER, R. B. Nonroman scripts and computer terminal developments. Information technology and libraries (Chicago), Vol. 1, No. 2, 1982, pp. 143-148.

335. MIRABAIL, M. Les 50 premiers mots clefs de la télématique. Toulouse, Privat, 1981, 314p.

336. MISHIN, V. A. Oprintsipakh i metodakh avtomaticheskoi reklassifikatsii. Sovetskoya bibliotekovedenie (Moscow), Vol. 3, 1979, pp. 56-68.

337. MLA handbook for writers of research papers, theses, and dissertations. New York, Modern Language Association of America, 1977, 163p.

338. MORIN-LABATUT, G.; SLY, M. Recommended methods for development information systems. Volume 1. Manual for the preparation of records in development information systems. Ottawa, International Development Research Centre, 1982, 270p. (approximate)

339. MORITA, I. Japanese character input: its state and problems. Journal of library automation (Chicago), Vol. 14, No. 1, 1981, pp. 6-23.

340. MORRISSEY, J. An 'intelligent terminal' for implementing relevance feedback on large operational retrieval systems. In: SALTON, G.; SCHNEIDER, H.-J., eds. Research and development in information retrieval, proceedings, Berlin, May 1982. Berlin, Springer-Verlag, 1983, pp. 38-50.

341. MORTON, S. E. Designing a multilingual terminology data bank for United States translators. Journal of the American Society for Information Science (New York), Vol. 29, No. 6, 1978, pp. 297-303.

342. MOULINOUX, C.; FAURE, J. C.; LITWIN, W. MESSIDOR: a distributed information retrieval system. In: SALTON, G.; SCHNEIDER, H.-J., eds. Research and development in information retrieval, proceedings, Berlin, May 1982. Berlin, Springer-Verlag, 1983, pp. 51-61.

343. MUENCH, E. V. A computerized English-Spanish correlation index to five biomedical library classification schemes based on MeSH. Bulletin of the Medical Library Association (Chicago), Vol. 59, No. 3, 1971, pp. 404-411.

344. MULDOWNEY, B. Buying plug-compatible tape and disk drives. Computer Decisions (Rochelle Park), Vol. 10, No. 3, 1978, pp. 40-44.

345. MULVIHILL, J. G. Impact of international data exchange on GeoRef. In: KIDD, C. M., ed. Proceedings of the Second International Conference of Geological Information. Norman, Oklahoma Geological Survey, [in press].

346. MURDOCK, J. W. Numeric data indexing. Journal of chemical information and computer sciences (Washington, DC), Vol. 20, No. 3, 1980, pp. 132-136.

347. NAHOURAII, E.; BROOKS, L. O.; CARDENAS, A. F. An approach to data communication between different generalized data base management systems. In: LOCKEMANN, P. C.; NEUHOLD, E. J., eds. Systems for large data bases: proceedings of the 2nd international conference on very large data bases, Brussels, 8-10 September 1976. Amsterdam, North-Holland, 1977, pp. 117-142.

348. NARIN, F.; CARPENTER, M. P. National publication and citation comparisons. Journal of the American Society for Information Science (Washington, DC), Vol. 26, No. 2, 1975, pp. 80-93.

349. NASH, M. Videotex versus online bibliographic databases: a Canadian perspective. Online review (Oxford), Vol. 6, No. 4, 1982, pp. 291-296.

350. NATIONAL FEDERATION OF SCIENCE ABSTRACTING AND INDEXING SERVICES. Data element definitions for secondary services. In: KEENAN, S., ed. Key papers on the use of computer-based bibliographic services. Washington, DC, American Society for Information Science, 1973, pp. 71-79.

351. NCLIS/NBS TASK FORCE ON COMPUTER NETWORK PROTOCOL. A computer network protocol for library and information science applications. Washington, DC, National Commission on Libraries and Information Science, 1977, VII, 90p.

352. NEALE, W. E. RR1--micrographic standards and related items. Journal of micrographics (Silver Spring, MD), Vol. 14, No. 6, 1981, pp. 35-40.

353. NEARY, D. R. Select an information storage system compatible with future technologies. Journal of micrographics (Silver Spring, MD), Vol. 15, No. 9, 1982, pp. 15-18.

354. NECAS, J. Input-output devices. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 666-707.

355. NECAS, J. Keyboard standards. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 747-749.

356. NEELAMEGHAN, A. Absolute syntax and structure of an indexing and switching language. In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 165-176.

357. NEGUS, A. E. Development of the EURONET-DIANE common command language. In: 3rd international online information meeting, London, 4-6 December 1979. Oxford, Learned Information, 1979, pp. 95-98.

358. NEGUS, A. E. EURONET guideline: standard commands for retrieval systems (Final report on a study carried out for the Commission of the European Communities, DG XIII). London, INSPEC, Institution of Electrical Engineers, 1977, 66p.

359. NEGUS, A. E. Extension of the Euronet guideline: standard commands for retrieval systems. Study final report. London, Scicon Consultancy International, 1980, 38p.

360. NEGUS, A. E. Investigation into the feasibility of the outside operation of EURODICAUTOM. Bedfordshire, 1981, 70p.

361. NEGUS, A. E. Why standards for command languages? In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 425-431.

362. NEGUS, A. E.; SNOWDEN, A. E. EURONET-DIANE user's guide (common command language). London, Scicon Consultancy International, 1980, 30p.

363. NEICU, S. D. Languages, thesauri, standardization. Probleme de informare si documentare (Bucarest), Vol. 9, No. 6, 1975, pp. 348-364.

364. NEUMANN, A. J. A guide to networking terminology. Information reports & bibliographies (New York), Vol. 7, No. 1, 1978, pp. 15-24.

365. NEUSTADT, R. M The birth of electronic publishing: legal and economic issues in telephone, cable and over-the-air teletext and videotext. White Plains, NY, Knowledge Industry Publications, 1982, III, 145p.

366. NEVILLE, H. H. Feasibility study of a scheme for reconciling thesauri covering a common subject. Journal of documentation (London), Vol. 26, No. 4, 1970, pp. 313-336.

367. NEWELL, A.; SPROULL, R. F. Computer networks: prospects for scientists. Science (Washington, DC), Vol. 215, No. 4534, 1982, pp. 843-852.

368. NIEHOFF, R. T. Development of an integrated energy vocabulary and the possibility for on-line subject switching. Journal of the American Society for Information Science (Washington, DC), Vol. 27, No. 1, 1976, pp. 3-17.

369. NIEHOFF, R. T. The optimization and use of automated subject switching for better retrieval. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 17, 1980, pp. 397-400.

370. NIEHOFF, R. T.; KWASNY, S.; WALKER, A.; WESSELLS, M. The design and evaluation of a vocabulary switching system for use in multi-base search environments. Columbus, Battelle Columbus Laboratories, 1980, 191p.

371. NIEHOFF, R. T.; KWASNY, S.; WESSELLS, M. Overcoming the data base vocabulary barrier--a solution. Online (Weston, Conn.), Vol. 3, No. 4, 1979, pp. 43-54.

372. NOLAN, R. E.; WHITNEY, V. K. M. Data element standardization: an approach with forward compatibility. Draft final report. Phase 1. Vienna, Va., Evaluation Research Corporation, 1978, II, 46p.

373. NOLAN, R. E.; WHITNEY, V. K. M. Data element standardization: a methodology for describing numeric data. In: GLAESER, P. S., ed. Data for science and technology: Proceedings of the Seventh International CODATA Conference, Kyoto, Japan, 8-11 October 1980. Oxford, Pergamon Press, 1981, pp. 535-539.

374. NORMAN, A. Electronic document delivery: the ARTEMIS concept for document digitalisation and teletransmission. White Plains, NY, Knowledge Industry Publications, 1982, XIV, 226p.

375. ODEGARD, W. M. Universally shared packet switched data networks. In: Sharing resources--sharing costs; proceedings of the seventh Canadian conference on information science, Banff, 12-15 May 1979. Ottawa, Canadian Association for Information Science, 1979, pp. 225-234.

376. ORI INC. Development of the WEECN vocabulary guide. Bethesda, Maryland, ORI Inc., Information Systems Division, 1979, 218p.

377. Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977. Munich, Verlag Dokumentation, 1977. 2 vols.

378. PACKER, A.; CONNING, A. M. The utilization of the UNISIST reference manual in the DOCPAL system: a DOCPAL-ISIS to UNISIST format conversion programme. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April 1978. Budapest, OMKDK-Technoinform, 1978, pp. 25-39.

379. PAINTER, A. F. An analysis of duplication and consistency of subject indexing involved in report handling at the Office of Technical Services, U.S. Department of Commerce. New Brunswick, New Jersey, Rutgers, the State University, 1963, IX, 135p. (PB 181501)

380. PARÉ, M. Les banques automatisées de terminologies multilingues et les organismes de normalisation. In: International cooperation in technology: first Infoterm symposium. Munich, Verlag Dokumentation, 1976, pp.224-233.

381. PARIS, J.; BOSS, R. W. The care and maintenance of videodiscs and players. Videodisc/videotex (Westport, Conn.), Vol. 2, No. 1, 1982, pp. 38-46.

382. PARK, M. K. Bibliographic and information processing standards. Annual review of information science and technology (White Plains, NY), Vol. 12, 1977, pp. 59-80.

383. PARK, M. K.; STEARNS, R. E., JR. Evaluation of the IMD IV+V-system software package for UNISIST reference manual implementations. Athens, GA, University of Georgia, 1982, 16p.

384. PATERSON, D. G.; TINKER, M. A. Readability of newspaper headlines printed in capitals and lower case. *Journal of applied psychology* (Lancaster, PA), Vol. 30, No. 1, 1946, pp. 161-168.

385. PENNIMAN, W. D.; KROHN, R. E.; KOVACS, G. J. A framework for the study of emerging network technology. *Journal of the American Society for Information Science* (Washington, DC), Vol. 25, No. 6, 1974, pp. 378-380.

386. PERREAULT, J. M. Some problems in the BSO. *International forum on information and documentation* (Moscow), Vol. 4, No. 3, 1979, pp. 16-20.

387. PERSHIKOV, V. F. The facet-block principle in designing an IRL for an automated IRS. *Automatic documentation and mathematical linguistics* (New York), Vol. 11, No. 4, 1977, pp. 20-27. [English translation of *Nauchno-tekhnicheskaya informatsiya, seriya 2* (Moscow), Vol. 11, No. 10, 1977, pp. [illegible]-20.]

388. PIGUR, V. A. The linguistic basis for the construction of multilingual thesauruses: selecting equivalents. *Automatic documentation and mathematical linguistics* (New York), Vol. 13, No. 6, 1979, pp. 19-31. [English translation of *Nauchno-tekhnicheskaya informatsiya, seriya 2* (Moscow), Vol. 13, No. 11, 1979, pp. 11-18.]

389. PITERNICK, A. B. *Bibliographic description in library cataloguing and in abstracting and indexing services with special emphasis on services in science and technology*. Vancouver, University of British Columbia, School of Librarianship, 1979, 64p.

390. PITERNICK, A. B. Standardization of journal titles in data bases. *Journal of the American Society for Information Science* (New York), Vol. 33, No. 2, 1982, p. 105.

391. PLANTE, S.; GRUSON, C.; WOLFF-TERROINE, M. Le macrothesaurus des sciences et techniques. *Documentaliste* (Paris), Vol. 14, No. 1, 1977, pp. 20-26.

392. POBUKOVSKY, M. Unesco's integrated documentation network: the CDS/ISIS system. *Unesco journal of information science, librarianship and archives administration* (Paris), Vol. 2, No. 3, 1980, pp. 195-20[illegible]

393. POLLITT, A. S. An expert system as an online search intermediary. *In: 5th international online information meeting, London, 8-10 December 1981*. Oxford, Learned Information, 1981, pp. 25-32.

394. PONCET, J. Authority files in machine systems. *In*: COWARD, R. E.; YELLAND, M., *eds*. *The interchange of bibliographic information in machine readable form*. London, Library Association, 1975, pp. 96-98.

395. POUZIN, L. Some data communications gaps. *In*: CSABA, L.; SZENTIVÁNYI, T.; TARNAY, K., *eds*. *Networks from the user's point of view* (proceedings of the IFIP TC-6 working conference COMNET '81, Budapest, 11-15 May 1981). Amsterdam, North-Holland, 1981, pp. 569-574.

396. POUZIN, L.; ZIMMERMANN, H. A tutorial on protocols. *IEEE proceedings* (New York), Vol. 66, No. 11, 1978, pp. 1346-1370.

397. PRICE, D. J. DE SOLLA. Little science, big science. New York, Columbia University Press, 1963, 119p.

398. PROTAPOVA, R. D. Issledovanie sovmestimosti tsentralizovannoi predmetizatsii s predmetizatsiei v otdel' nykh bibliotekakh. Sovetskaya bibliografia (Moscow), Vol. 153, 1975, pp. 22-30.

399. PURSER, M. The Euronet Diane network for information retrieval. Information technology: research and development (Guildford), Vol. 1, No. 3, 1982, pp. 197-216.

400. RAGHAVENDRA RAO, M. K.; KAMATH, V. A. INIS--a successful experiment in operating a decentralized, computer-based, mission-oriented information system. In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 454-465.

401. RAITT, D. Recent developments in telecommunications and their impact on information services. Aslib proceedings (London), Vol. 34, No. 1, 1982, pp. 54-76.

402. RANGANATHAN, S. R.; NEELAMEGHAN, A. Basis for study of compatibility, and of compatibility of the Colon Classification with the Universal Decimal Classification. Library science with a slant to documentation (Bangalore), Vol. 7, No. 2, 1970, pp. 97-109.

403. RASSAM, G.; GRAVESTEIJN, J.; LAGARDE, E. A joint international bibliographic data base in geology. Bulletin du B.R.G.M. (Paris), Section IV, No. 3, 1980/81, pp. 215-221.

404. RATHER, J. C. The realities of interchanging machine-readable bibliographic records. In: The exchange of bibliographic data and the MARC format. Munich, Verlag Dokumentation, 1972, pp. 27-35.

405. RATHER, L. J.; DE LA GARZA, P. J. Getting it all together: international cataloging cooperation and networks. Journal of library automation (Chicago), Vol. 10, No. 2, 1977, pp. 163-169.

406. Reference manual for machine-readable bibliographic descriptions. Paris, UNESCO, 1974, 71p. (Second edition, 1981, in looseleaf form)

407. RIGG, J. C. Presentation of numeric data in publications. Journal of research communication studies (Amsterdam), Vol. 3, No. 4, 1982, pp. 367-376.

408. RIGG, P. The new data networks--now and in the near future. Program (London), Vol. 14, No. 2, 1980, pp. 62-68.

409. RIGGS, F. W. A new paradigm for social science terminology. International classification (Munich), Vol. 6, No. 3, 1979, pp. 150-158.

410. ROGERS, F. B.; CHAREN, T. Abbreviations for medical journal titles. Bulletin of the Medical Library Association (Chicago), Vol. 50, No. 3, 1962, pp. 311-350.

411. ROLLING, L. Compilation of thesauri for use in computer systems. Information storage and retrieval (Elmsford, U.K.), Vol. 6, No. 4, 1970, pp. 341-350.

412. ROLLING, L. Computer management of multi-lingual thesauri. In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 382-388.

413. ROLLING, L. Graphic display devices in thesaurus construction and use. Aslib proceedings (London), Vol. 23, No. 11, 1971, pp. 591-594.

414. ROLLING, L. The second birth of machine translation: a timely event for data base suppliers and users. Electronic publishing review (Oxford), Vol. 1, No. 3, 1981, pp. 211-216.

415. ROSENTHAL, R. A review of network access techniques with a case study: the Network Access Machine. Washington, DC, Institute for Computer Sciences and Technology, National Bureau of Standards, 1976, IV, 29p. (NBS technical note 917)

416. ROSENTHAL, R.; LUCAS, B. D. The design and implementation of the National Bureau of Standards' Network Access Machine (NAM). Washington, DC, Institute for Computer Sciences and Technology, National Bureau of Standards, 1978, IV, 43p. (NBS special publication 500-35)

417. ROSS, H. M. Character sets for communication of text including a suggestion for a Unified Coding Scheme for the text communication service (prepared for the Commission of the European Communities Directorate General for Scientific and Technical Information and Information Management). London, Data Systems Consultants, 1979, 92p. + appendices.

418. ROSSMASSLER, S. A.; WATSON, D. G., eds. Data handling for science and technology: an overview and sourcebook. Amsterdam, North-Holland, 1980, XVI, 184p.

419. ROUSE, S. H.; LANNOM, L. W. Some differences between three on-line systems: impact on search results. Online review (Oxford), Vol. 1, No. 2, 1977, pp. 117-132.

420. ROWLETT, R. J.; WIGINGTON, R. L.; WOOD, J. L. Standard distribution format: Chemical Abstracts Service. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April 1978. Budapest, OMKDK-Technoinform, 1978, pp. 105-111.

421. RUBIN, M. L., ed. Documentation standards and procedures for online systems. New York, Van Nostrand Reinhold, 1979, XV, 251p.

422. SAAL, H. J. Local-area networks: possibilities for personal computers. Byte (Peterborough, NH), Vol. 6, No. 10, 1981, pp. 92-112.

423. SAFFADY, W. Micrographics. Littleton, Colo., Libraries Unlimited, 1978, 238p.

424. SAFFADY, W. Software: a tutorial introduction. Software review (Westport, Conn.), Vol. 1, No. 1, 1982, pp. 5-10.

425. SAGER, J. C.; SOMERS, H. L.; MCNAUGHT, J. Thesaurus integration in the social sciences. Part 1. Comparison of thesauri. International classification (Munich), Vol. 8, No. 3, 1981, pp. 133-138.

426. SAHLI, N. SPINDEX users network. In: MCCRANK, L. J., ed. Automating the archives: issues and problems in computer applications. White Plains, NY, Knowledge Industry Publications, 1981, pp. 93-121.

427. SALISBURY, B. A., JR.; STILES, H. E. The use of the B-coefficient in information retrieval. Proceedings of the American Society for Information Science, Vol. 6, 1969, pp. 265-268.

428. SALTON, G. Automatic processing of foreign language documents. Journal of the American Society for Information Science (Washington, DC), Vol. 21, No. 3, 1970, pp. 187-194.

429. SALTON, G.; MCGILL, M. J. Introduction to modern information retrieval. New York, McGraw-Hill, 1983, XV, 448p.

430. SAMMET, J. E. Programming languages: history and fundamentals. Englewood Cliffs, NJ, Prentice-Hall, 1969, XXX, 785p.

431. SAUER, M. E. Key title and rules for entry. Library resources and technical services (Chicago), Vol. 19, No. 4, Fall 1975, pp. 338-340.

432. SCHARFF, L. Euronet-DIANE: how to achieve further harmonization. In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 417-423.

433. SCHIPMA, P. B.; WILLIAMS, M. E.; SHAFTON, A. L. Comparison of document data bases. Journal of the American Society for Information Science (Washington, DC), Vol. 22, No. 5, 1971, pp. 326-332.

434. SCHULMAN, J.-L. Video PATSEARCH: a mixed-media system. Information technology and libraries (Chicago), Vol. 1, No. 2, 1982, pp. 150-156.

435. SCHULTHEISZ, R. J. TOXLINE: evolution of an online interactive bibliographic data base. Journal of the American Society for Information Science (New York), Vol. 32, No. 6, 1981, pp. 421-429.

436. SCHULZ, J. A terminology data bank for translators: methods of interrogation in the TEAM system. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 153-188.

437. SEBESTYEN, I. The videodisc revolution. Electronic publishing review (Oxford), Vol. 2, No. 1, 1982, pp. 41-89.

438. SEDMAN, E. C. The use of MicroCobol for telesoftware. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 399-411.

439. SEMENOV, V. F.; VASIL'EVA, N. V.; BYKOVA, L. N.; KIPTENKO, V. A.; STREL'TSOVA, T. V. Creation of a bilingual thesaurus for indexing documents and inquiries in the field of nuclear science and technology. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 167-174.

440. SHEMAKIN, YU. I; KULIK, A. N. Compatibility of information retrieval systems and the function of a polytechnical thesaurus in a unified information network. Automatic documentation and mathematical linguistics (New York), Vol. 4, No. 3, pp. 35-41. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2, No. 8, 1970, pp. 30-34.]

441. SHERWOOD, B. A. Computer diacritics. Urbana-Champaign, University of Illinois, Computer-based Education Research Laboratory, 1982, 6p.

442. SHILLING, C. W. The need for compatibility in biological science. In: NEWMAN, S. M., ed. Information systems compatibility. Washington, DC, Spartan Books, 1965, pp. 3-5.

443. SHNEIDERMAN, B. Data type. In: RALSTON, A.; MEEK, C. L., eds. Encyclopedia of computer science. New York, Petrocelli/Charter, 1976, pp. 442-443.

444. SHNEIDERMAN, B. Software psychology: human factors in computer and information systems. Cambridge, Mass., Winthrop Publishers, 1980, XV, 320p.

445. SIBLEY, E. H. Standardization and database systems. ACM SIGMOD record (New York), Vol. 9, No. 4, 1977, pp. 144-155.

446. SIMON, R. C.; DOSZKOCS, T. E. Implementation of a user interface for a complex numeric/textual data base. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 15, 1978, pp. 309-313.

447. SINGLETON, W. T. Display design: principles and procedures. IEEE transactions on man-machine systems (New York), Vol. MMS-10, No. 4, 1969, pp. 181-193.

448. SKILLIN, M. E.; GAY, R. M. Words into type. Third edition. Englewood Cliffs, NJ, Prentice-Hall, 1974, 585p.

449. SMIRLE, J. C.; LUM, Y. F.; BROWN, H. G. International videotex standardization: a Canadian view of progress towards the wired world. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 271-280.

450. SMITH, F. G. Data communications and the systems designer. Ann Arbor, UMI Research Press, 1980, IX, 201p.

451. SMITH, L. C. Systematic searching of abstracts and indexes in interdisciplinary areas. Journal of the American Society for Information Science (Washington, DC), Vol. 25, No. 6, 1974, pp. 343-353.

452. SOBOLEVSKAYA, E. Z. Sravnitel' nyi anliz predmetizatsii dokumentov osyshchestulyaemoi vsesoyuznoi knizhnoi palatoi i gosudarstvennoi publichnoi bibliotekoi im M.E. Saltykova-Shchedrina. Nauchnye i tekhnicheskie biblioteki SSSR (Moscow), Vol. 7, 1976, pp. 10-16.

453. SOERGEL, D. The Broad System of Ordering: a critique. International forum on information and documentation (Moscow), Vol. 4, No. 3, 1979, pp. 21-24.

454. SOERGEL, D. A general model for indexing languages: the basis for compatibility and integration. In: WELLISCH, H.; WILSON, T. D., eds. Subject retrieval in the seventies: new directions. Westport, Conn., Greenwood Publishing Co., 1972, pp. 36-61.

455. SOERGEL, D. Indexing languages and thesauri: construction and maintenance. Los Angeles, Melville Publishing Co. (John Wiley), 1974, XIII, 632p.

456. SOKOLOV, A. V. A way to ensure thesaurus compatibility. Automatic documentation and mathematical linguistics (New York), Vol. 11, No. 1, 1977, pp. 22-30. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 11, No. 1, 1977, pp. 19-24.]

457. SOKOLOV, A. V.; MANKEVICH, I. A.; FEDOTOVA, A. P. An experiment in the development of compatible specialized information-retrieval thesauri. Automatic documentation and mathematical linguistics (New York), Vol. 12, No. 4, 1978, pp. 8-19. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 12, No. 10, 1978, pp. 5-12.]

458. SOMERFIELD, G. A. The secondary services: the use of computers in the production of abstract and indexing services. In: BIDMEAD, M., ed. Use of computers in libraries and information centres. London, Aslib, 1976, pp. 66-82.

459. SOROKIN, YU. N. Principles involved in creating an international scientific and technical information system for the member states of the Council of Mutual Economic Assistance (CMEA). In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 257-264.

460. SPARCK JONES, K. Automatic indexing 1974: A state of the art review. Cambridge, England, University of Cambridge, 1974, 85p.

461. SPINRAD, R. J. Office automation. Science (Washington, DC), Vol. 215, No. 4534, 1982, pp. 808-813.

462. Standard practices in the preparation of bibliographic records. London, IFLA International Office for UBC, 1982, X, 47p. (Occasional paper No. 9)

463. STEVENS, M. E. Automatic indexing: A state-of-the-art report. Washington, DC, National Bureau of Standards, 1980, V, 220p.

464. STEVENS, M. E. Compatibility problems of network interfacing. In: BECKER, J., ed. Proceedings of the conference on interlibrary communications and information networks, Airlie House, Warrenton, Virginia, 28 September - 2 October 1970. Chicago, American Library Association, 1971, pp. 202-212.

465. STEWART, A. K. Selection and use of equipment to manipulate search output. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 18, 1981, pp. 251-252.

466. STIBIC, V. A few practical remarks on the user-friendliness of online systems. Journal of information science (Amsterdam), Vol. 2, No. 6, 1980, pp. 277-283.

467. STILES, H. E. Machine retrieval using the association factor. In: Machine indexing, progress and problems. Washington, DC, American University, 1961, pp. 192-206.

468. STOKELY, E. E. Standardization in information systems: some pros and cons. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 349-351.

469. STOREY, I. N. J. The future of videotex systems--a personal view. Journal of information science (Amsterdam), Vol. 4, No. 2/3, 1982, pp. 111-115.

470. SUTHASINEKUL, S. Microfilm vs. optical disc as storage medium for document retrieval and dissemination. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 17, 1980, pp. 100-102.

471. SYROVÝ, K. Současné podmínsky pro uplatnění automatizace v soustavě VTEI. Ceskoslovenska informatika (Prague), Vol. 20, No. 2, 1978, pp. 38-41.

472. SZILVASSY, J. ISDA: worldwide serials control. IFLA journal (Munich), Vol. 8, No. 4, 1982, pp. 371-378.

473. TAKANO, F. The online information service in JICST and the special characteristics of the Japanese language: implementation of automatic processing of Japanese. In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 73-90.

474. TANENBAUM, A. S. Network protocols. ACM computing surveys (New York), Vol. 13, No. 4, 1981, pp. 453-489.

475. TANTAWI, A. N.; MOURAD, M. M. Workstations in the electronic office. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 159-171.

476. TEDD, L. Intelligence in the user's terminal: a look at current options and possibilities. In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 1-10.

477. TENNANT, H. Natural language processing: an introduction to an emerging technology. New York, Petrocelli, 1981, VIII, 275p.

478. TENNE-SENS, A. Telidon graphics and library applications. Information technology and libraries (Chicago), Vol. 1, No. 2, 1982, pp. 98-110.

479. Terminological data banks. Munich, K.G. Saur, 1980, 206p. (Infoterm series 5)

480. Terminologies for the eighties. Munich, K.G. Saur, 1982, 412p. (Infoterm series 7)

481. Theoretical and methodological problems of terminology. Munich, K.G. Saur, 1981, 608p. (Infoterm series 6)

482. THOMA, G. R. Transmission of information: an overview. Journal of the American Society for Information Science (New York), Vol. 32, No. 2, 1981, pp. 131-140.

483. TOLIVER, D. OL'SAM: an intelligent front-end for bibliographic information retrieval. Information technology and libraries (Chicago), Vol. 1, No. 4, 1982, pp. 317-326.

484. TOLIVER, D. A program for machine-mediated searching. Information processing & management (Oxford), Vol. 17, No. 2, 1981, pp. 61-68.

485. TOMA, P. P. SYSTRAN as a multilingual machine translation system. In: Overcoming the language barrier: [proceedings of the] Third European Congress on Information Systems and Networks, Luxembourg, 3-6 May 1977, Vol. 1. Munich, Verlag Dokumentation, 1977, pp. 569-581.

486. TOMAN, J.; LLOYD, G. A. Introduction to the Subject-field Reference Code (SRC) or Broad System of Ordering (BSO) for UNISIST purposes. In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 321-324.

487. TRAVIS, I. L. Describing and indexing federal statistical data systems. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 18, 1981, pp. 344-346.

488. TRELEAVEN, P. C.; LIMA, I. G. Japan's fifth-generation computer systems. Computer (New York), Vol. 15, No. 8, 1982, pp. 79-88.

489. TRESSEL, G. W.; PENNIMAN, W. D. The effect of transmission techniques on the future of microforms. Drexel library quarterly (Philadelphia), Vol. 11, No. 4, 1975, pp. 75-82.

490. TURKOV, ZH.; CHEPKASOV, A. Planning and development of INIS: its application of UNISIST principles and its interconnection with other international and national systems. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 15-34.

491. TURNER, B. IAPS notebook. Interface age (Cerritos), Vol. 3, No. 11, 1978, pp. 74-79.

492. TURTLE, H.; PENNIMAN, W. D.; HICKEY, T. B. Data entry/display devices for interactive information retrieval. Annual review of information science and technology (White Plains, NY), Vol. 16, 1981, pp. 55-83.

493. TYNER, S. Checklist for printing terminals. Journal of library automation (Chicago), Vol. 13, No. 1, 1980, pp. 108-118.

494. UNDERWOOD, M. J. The present situation regarding vocal data acquisition. Information services & use (Amsterdam), Vol. 1, No. 2, 1981, pp. 57-64.

495. UNIMARC: universal MARC format. Second edition. London, IFLA International Office for UBC, 1980, XII, 131p.

496. UNISIST guide to standards for information handling; compiled by E. Vajda. Paris, Unesco, 1980, 304p.

497. *UNISIST guidelines for the establishment and development of multilingual thesauri*. Revised edition. Paris, Unesco, 1980, 85p. (PGI/80/WS/12)

498. *UNISIST: study report on the feasibility of a world science information system*. Paris, Unesco, 1971, IV, 161p.

499. UNITED NATIONS. Dag Hammarskjold Library. *Bibliographical style manual*. New York, 1963, 62p.

500. UNITED NATIONS. Inter-agency Technical Experts Working Group on Compatibility with PADIS. *Report*. Addis Ababa, 22-25 April 1981, 30p.

501. U.S. Federal Council for Science and Technology. *COSATI subject category list*. Washington, DC, 1964, V, 55p. (AD 612200)

502. U.S. Library of Congress. *Bibliographical procedures and style*. Washington, DC, Government Printing Office, 1954, 127p.

503. USSR STATE COMMITTEE FOR SCIENCE AND TECHNOLOGY. Interdepartmental Commission on Classification. Comments on the Broad System of Ordering. *International forum on information and documentation* (Moscow), Vol. 4, No. 3, 1979, pp. 25-27.

504. VALANTIN, R. L. *CDS/ISIS and MINISIS: a functional analysis and comparison*. Ottawa, International Development Research Centre, 1981, 88p. (IDRC-TS37e)

505. VANDERHEIDEN, G. Modifying and designing computer terminals to allow access by handicapped individuals. *In*: DIVILBISS, J. L., *ed*. *Public access to library automation* (papers presented at the 1980 Clinic on Library Applications of Data Processing, 20-23 April 1980). Urbana-Champaign, University of Illinois, Graduate School of Library and Information Science, 1981, pp. 99-116.

506. VAN DIJK, M. Un thesaurus multilingue au service de la coopération internationale. *Bulletin de l' Association Internationale des Documentalistes et Techniciens de l' Information* (Paris), Vol. 5, 1966, pp. 85-87.

507. VAN SLYPE, G. SYSTRAN: evaluation of the 1978 version of the SYSTRAN English-French automatic system of the Commission of the European Communities. *The incorporated linguist* (London), Vol. 18, No. 3, 1979, pp. 86-89.

508. VAN SLYPE, G.; LAUDET, E.; MACHGEELS, C.; VANDER AUWERA, Y. A mini-Euronet standard command set at Université Libre de Bruxelles (ULB). *Online review* (Oxford), Vol. 4, No. 1, 1980, pp. 41-53.

509. VASARHELYI, P. The relevance of INTERCONCEPT for classification and indexing. *International classification* (Munich), Vol. 7, No. 1, 1980, pp. 6-9.

510. VASKEVITCH, D. Between three worlds: a strategy for computer and terminal independence. *Proceedings of the American Society for Information Science* (White Plains, NY), Vol. 19, 1982, pp. 310-313.

511. VERHEIJEN-VOOGD, C. Is EURONET C.C.L. a 'common' command language? *Online review* (Oxford), Vol. 5, No. 5, 1981, pp. 399-401.

512. VERONA, E. A decade of IFLA's work on the standardization of bibliographic description. IFLA journal (Munich), Vol. 6, No. 3, 1980, pp. 216-233.

513. VERVLIET, H. D. L. ISDS, ISBD(S), and ISBD(M) as cataloging aids. Journal of library automation (Chicago), Vol. 9, No. 2, 1976, pp. 137-144.

514. VICKERY, B. C. The organization of information using computers. In: BIDMEAD, M., ed. Use of computers in libraries and information centres. London, Aslib, 1976, pp. 25-33.

515. VICKERY, B. C.; MCILWAINE, I. C. Structuring and switching: a discussion of the Broad System of Ordering. International forum on information and documentation (Moscow), Vol. 4, No. 3, 1979, pp. 13-15.

516. VILENSKAYA, S. K. Information retrieval language compatibility (the intermediary language and the switching language). Automatic documentation and mathematical linguistics (New York), Vol. 11, No. 2, 1977, pp. 54-62. [English translation of Nauchno-tekhnicheskaya informatsiya, seriya 2 (Moscow), Vol. 11, No. 5, 1977, pp. 16-21.]

517. VITUKHNOVSKAYA, A. A. K teorii sovmestimosti informatsionno-poiskovykh tezaurusov. Nauchnye i tekhnicheskie biblioteki SSSR (Moscow), Vol. 11, 1976, pp. 6-14.

518. VOLNY, J. Jednotný informační systém v zemích RVHP. Knižnice a vedecké informácie (Martin), Vol. 9, No. 4, 1977, pp. 163-166.

519. VOOS, H. Telecommunications and facsimile. Special libraries (New York), Vol. 72, No. 2, 1981, pp. 118-121.

520. WAINWRIGHT, J. Standards used by bibliographic tape services: a comparison. London, Aslib, 1974, 27p. (OSTI report 5191)

521. WALKER, M. G.; HARRISON, R. G. Program portability. Datamation (New York), Vol. 28, No. 1, 1982, pp. 140-149.

522. WALL, E.; BARNES, J. M. Intersystem compatibility and convertibility of subject vocabularies. Philadelphia, Auerbach Corporation, 1969, IV, 144p. (Technical report 1582-100-TR-5)

523. WANG, C.; DUNN, R. Chinese character computer encoding methods and library automation: a bibliographical survey. LC information bulletin (Washington, DC), Vol. 41, No. 13, 1982, pp. 95-100.

524. WANGER, J.; LANDAU, R. N. Nonbibliographic on-line data base services. Journal of the American Society for Information Science (New York), Vol. 31, No. 3, 1980, pp. 171-180.

525. WASHINGTON LIBRARY NETWORK; THE RESEARCH LIBRARIES GROUP, INC.; LIBRARY OF CONGRESS. Linked Systems Project. Proposal to the Council on Library Resources for a standard network interconnection, 1981, 52p.

526. WASSERMAN, A. I. Information system design methodology. Journal of the American Society for Information Science (New York), Vol. 31, No. 1, 1980, pp. 5-24.

527. WATSON, D. G. Accessibility and dissemination of data. In: ROSSMASSLER, S. A.; WATSON, D. G., eds. Data handling for science and technology: an overview and sourcebook. Amsterdam, North-Holland, 1980, pp. 131-139.

528. WEISS, S. Online bibliographic services: a comparison. Special libraries (New York), Vol. 72, No. 4, 1981, pp. 379-389.

529. WELLISCH, H. H. The arrangement of entries in non-Roman scripts in multiscript catalogs and bibliographies. International forum on information and documentation (Moscow), Vol. 3, No. 3, 1978, pp. 18-24.

530. WELLISCH, H. H. The Broad System of Ordering, or Bishop Wilkins redivivus: a review article. Library quarterly (Chicago), Vol. 49, No. 4, 1979, pp. 444-452.

531. WELLISCH, H. H. Capital games: the problem of compatibility of bibliographic citations in data bases and in printed publications. FIPS software documentation workshop, March 3, 1982, program and working papers. Gaithersburg, MD, National Bureau of Standards, 1982, pp. 213-216.

532. WELLISCH, H. H. The conversion of scripts: its nature, history and utilization. New York, Wiley, 1978, XVIII, 509p.

533. WELLISCH, H. H. The cybernetics of bibliographic control: toward a theory of document retrieval systems. Journal of the American Society for Information Science (New York), Vol. 31, No. 1, 1980, pp. 41-50.

534. WELLISCH, H. H. The exchange of bibliographic data in non-Roman scripts. In: DIERICKX, H.; HOPKINSON, A., eds. Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 40-46.

535. WELLISCH, H. H. Linguistic and semantic problems in the use of English-language information services in non-English-speaking countries. International library review (New York), Vol. 5, No. 2, 1973, pp. 147-162.

536. WELLISCH, H. H. Multiscript and multilingual bibliographic control: alternatives to romanization. Library resources and technical services (Chicago), Vol. 22, No. 2, 1978, pp. 179-189.

537. WERSIG, G. Experiences in compatibility research in documentary languages. In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 423-430.

538. WESTCOTT, B. A. Computer communications, today and tomorrow--a tutorial. In: Sharing resources--sharing costs; proceedings of the seventh Canadian conference on information science, Banff, 12-15 May 1979. Ottawa, Canadian Association for Information Science, 1979, pp. 203-215.

539. What do librarians want in a system? (Besides Macedonian characters, diacrits, cedillas and Hebrew). Data management (Park Ridge, IL), Vol. 20, No. 1, 1982, pp. 32-34.

540. WHEELER, J. L. International standards activities. In: Proceedings of computer networking symposium, Gaithersburg, MD, 15 December 1977. New York, Institute of Electrical and Electronics Engineers, 1977, pp. 26-35.

541. WHITEHEAD, J. B. Communications using word processing systems. Information services & use (Amsterdam), Vol. 1, No. 3, 1981, pp. 109-138.

542. WHITEHEAD, J. B. Word processing: an introduction and appraisal. Journal of documentation (London), Vol. 36, No. 4, 1980, pp. 313-341.

543. WILD, K.; WOOLSTON, J. E. A family of international information systems as a resource for economic and social development. Paper presented at the 38th FID World Congress, Mexico City, 1976.

544. WILLIAMS, M. E. Online retrieval--today and tomorrow. Online review (Oxford), Vol. 2, No. 4, 1978, pp. 353-366.

545. WILLIAMS, M. E.; LANNOM, L. Lack of standardization of the journal title data element in databases. Journal of the American Society for Information Science (New York), Vol. 32, No. 3, 1981, pp. 229-233.

546. WILLIAMS, M. E.; PREECE, S. E. Automatic data base selector for network use. Proceedings of the American Society for Information Science (White Plains, NY), Vol. 14, 1977, p. 34. (Full paper on microfiche)

547. WILLIAMS, M. E.; SCHIPMA, P. B. Design and operation of a computer search center for chemical information. Journal of chemical documentation (Washington, DC), Vol. 10, No. 3, 1970, pp. 158-162.

548. WILLIAMS, P. W. A new device to simplify online searching and reduce costs. In: WILLIAMS, M. E.; HOGAN, T. H., comps. National online meeting proceedings, New York, 24-26 March 1981. Medford, N.J., Learned Information, 1981, pp. 503-514.

549. WILLIAMS, P. W. A proposal for a network computer for online assistance. In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 11-22.

550. WILLIAMS, P. W. The use of microcomputers for online computer access. In: WILLIAMS, M. E.; HOGAN, T. H., comps. National online meeting proceedings, New York, 31 March - 1 April 1982. Medford, N.J., Learned Information, 1982, pp. 559-568.

551. WILMOT, C. E. Economic aspects of a common bibliographic exchange format. In: Toward a common bibliographic exchange format? Proceedings of the International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978. Budapest, OMKDK-Technoinform, 1978, pp. 13-17.

552. WILMOT, C. E. International tape standards for documentation and information processing. London, Aslib, 1977, IV, 26p.

553. WILSON, P. Two kinds of power: an essay in bibliographic control. Berkeley, University of California Press, 1968, 155p.

554. WINTER, R.; LOZANO-PEREZ, T.; EASTLAKE, D. E. III; ROSENBERG, R. L. Overview of the Chemical Substances Information Network. Cambridge, Mass., Computer Corporation of America, 1981, 15p. (Technical Report CCA-81-03)

555. WITHINGTON, F. G. Computer technology: state of the art. Journal of the American Society for Information Science (New York), Vol. 32, No. 2, 1981, pp. 124-130.

556. WOLF, D. Impact of ISDS on the registration of serials in the Deutsche Bibliothek. In: HELAL, A. H.; WEISS, J. W., eds. Current trends in serials automation. Essen, Essen University Library, 1981, pp. 13-24.

557. WOLFF-TERROINE, M. A macrothesaurus. Why? How? In: NEELAMEGHAN, A., ed. Ordering systems for global information networks. Bangalore, International Federation for Documentation, 1979, pp. 431-436.

558. WOLFF-TERROINE, M. Multilingual systems. In: Second European Congress on Information Systems and Networks. Munich, Verlag Dokumentation, 1976, pp. 149-158.

559. WOODS, R. The compatibility of library systems. Journal of library automation (Chicago), Vol. 13, No. 4, 1980, pp. 244-250.

560. WOOLFE, R. Private videotex systems and gateway link to public services. In: 5th international online information meeting, London, 8-10 December 1981. Oxford, Learned Information, 1981, pp. 193-198.

561. WYSOCKI, A. UNISIST programme: its role in international co-operation in scientific and technical information. In: Information systems: their interconnection and compatibility. Vienna, International Atomic Energy Agency, 1975, pp. 3-13.

562. YASAKI, E. K. Tokyo looks to the '90s. Datamation (New York), Vol. 28, No. 1, 1982, pp. 110-115.

563. ZAMORA, A. Automatic detection and correction of spelling errors in a large data base. Journal of the American Society for Information Science (New York), Vol. 31, No. 1, 1980, pp. 51-57.

564. ZAYE, D. F. Data-tagging experiment for energy-related documents at CAS. Journal of chemical information and computer sciences (Washington, DC), Vol. 21, No. 2, 1981, pp. 73-78.

565. ZIMMERMANN, H.; NAFFAH, N. Protocols and network interconnection. In: BEAUCHAMP, K. G., ed. Interlinking of computer networks; proceedings of the NATO Advanced Study Institute held at Bonas, France, 28 August - 8 September 1978. Dordrecht, D. Reidel, 1979, pp. 261-272.

566. ZIMMERMANN, R. Future utilization of interactive and broadcast videotex in Germany and its effects on standardization. In: Viewdata and videotext, 1980-81: a worldwide report (transcript of viewdata '80, first world conference on viewdata, videotex, and teletext). White Plains, NY, Knowledge Industry Publications, 1980, pp. 263-269.

ANNEX: LIST OF ABBREVIATIONS AND ACRONYMS

AACR	Anglo-American Cataloguing Rules
AACR2	The second edition of AACR
AGROVOC	Agricultural vocabulary developed by FAO
AGI	American Geological Institute
AGRIS	International Information System for Agricultural Sciences and Technology
ALA	American Library Association
ANSI	American National Standards Institute
ARTEMIS	Automatic Retrieval of Text from Europe's Multinational Information Service
ASCII	American Standard Code for Information Interchange
ASIDIC	Association of Information and Dissemination Centers
ASTM	American Society for Testing and Materials
AUSTIC	All-Union Scientific and Technical Information Centre
BIOSIS	BioSciences Information Service of Biological Abstracts
BITS	BIOSIS Information Transfer System
BRGM	Bureau de Recherches Géologiques et Minières
BSI	British Standards Institution
BSO	Broad System of Ordering
CARISPLAN	Caribbean Information System for Economic and Social Planning
CAST	Chemical Automated Search Terminal
CCA	Computer Corporation of America
CCF	Common Communication Format
CCG	Computer Communications Group
CCITT	Comité Consultatif International Télégraphique et Téléphonique
CCL	Common Command Language
CDIC	Caribbean Documentation and Information Centre
CEC	Commission of the European Communities
CED	Centro Elettronico di Documentazione Giuridica della Corte Suprema di Cassazione
CELADE	United Nations Latin American Demographic Centre
CEPAL	Comisión Económica para América Latina (i.e., ECLA)
CEPT	Conférence Européene des Administrations des Postes et Télécommunications
CEQ	Council on Environmental Quality
CICIS	Chemicals In Commerce Information System
CILEA	Consorzio Interuniversitario Lombardo per la Elaborazione Automatica
CIP	Cataloguing-in-Publication
CIS	Chemical Information System
CIT	CSIN Intelligent Terminal
CITE	Current Information Transfer into English
CLADES	Centro Latinoamericano de Documentación Económica y Social
CMEA	Council for Mutual Economic Assistance
CNRS	Centre National de la Recherche Scientifique
CODATA	Committee on Data for Science and Technology (ICSU)
COM	Computer-Output Microform
CONIT	Conversion for Network Information Transfer
COSATI	Committee on Scientific and Technical Information
CPL	CSIN Pattern Language
CRGS	Chemical Regulations and Guidelines System
CRT	Cathode Ray Tube

CSIL	CSIN Script Implementation Language
CSIN	Chemical Substances Information Network
DBMS	Data Base Management System
DCE	Data Circuit Terminating Equipment
DDC	Dewey Decimal Classification (or Defense Documentation Center)
DDP	Distributed Data Processing
DEVSIS	The International Information System for the Development Sciences
DIANE	Direct Information Access Network for Europe
DILOS	Dialogue Information Logic System
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
DIN	Deutsches Institut für Normung
DIS	The Development Information System
DOCPAL	Latin American Population Documentation System
DRCS	Dynamically Refinable Character Set
DTE	Data Terminal Equipment
EBCDIC	Extended Binary Coded Decimal Interchange Code
ECA	Economic Commission for Africa
ECHO	European Commission Host Organization
ECLA	Economic Commission for Latin America
ECWA	Economic Commission for Western Asia
EIA	Electronic Industries Association
EPA	Environmental Protection Agency
ERIC	Educational Resources Information Center
ESA	Department of Economic and Social Affairs
ESA/IRS	European Space Agency/Information Retrieval Service
ESCAP	Economic and Social Commission for Asia and the Pacific
EUDISED	European Documentation and Information System for Education
EURONET	European On-line Information Network
FAO	Food and Agriculture Organization of the United Nations
FIZ-TECHNIK	FachInformationsZentrum Technik
FRED	Front End for Databases
GID	Gesellschaft für Information und Dokumentation
IAPS	International ASCII Publishing Standard
ICC	International Computing Centre
ICSTI	International Centre for Scientific and Technical Information (USSR)
ICSU	International Council of Scientific Unions
ICSU-AB	The Abstracting Board of ICSU
IDRC	International Development Research Centre
IESA	Department of International Economic and Social Affairs
IFLA	International Federation of Library Associations and Institutions
IIASA	International Institute for Applied Systems Analysis
IIDA	Individualized Instruction for Data Access
ILO	International Labour Organization
ILPES	Instituto Latinoamericano de Planificación Económica y Social
IMP	Interface Message Processor
INFOTERM	International Information Centre for Terminology
INIS	International Nuclear Information System
INKA	Informationssystem Karlsruhe
INSIS	Inter-institutional Integrated Services Information System
INSPEC	Information Service for Physics, Electrotechnology and Control
INTERMARC	International MARC Format

IOB	Inter-Organization Board for Information Systems
ISAN	International Standard Authority Number
ISBD	International Standard Bibliographic Description
ISBD(G)	A general ISBD framework
ISBD(M)	The ISBD for monographs
ISBD(S)	The ISBD for serials
ISBN	International Standard Book Number
ISCF	International Standard Control Form
ISCME	International Steering Committee for Medical Editors
ISDS	International Serials Data System
ISIS	Integrated Set of Information Systems
ISO	International Organization for Standardization
ISONET	ISO Information Network
ISSN	International Standard Serial Number
ISTIS	International Scientific and Technical Information System
ISTRN	International Standard Technical Report Number Code
ISU	The Information Systems Unit of IESA
ITSDC	Interagency Toxic Substances Data Committee
IUB	International Union of Biochemistry
IUNS	International Union of Nutritional Sciences
IUPAC	International Union of Pure and Applied Chemistry
IUPAP	International Union of Pure and Applied Physics
IWGDE	Interlaboratory Working Group for Data Exchange
JICST	Japan Information Center of Science and Technology
JSBD	Japanese Standard Bibliographic Description
KDEM	Kurzweil Data Entry Machine
LAN	Local Area Network
LC	Library of Congress
LCCN	Library of Congress Card Number
LCS	Library Control System
MAB-1	Maschinelles Austauschformat für Bibliotheken Version 1
MADOK	Magnetband-Austauschformat für Dokumentationzwecke
MARBI	Interdivisional Committee for the Representation in Machine-Readable Form of Bibliographic Information (ALA)
MARC	Machine-Readable Cataloging
MARCAL	The Latin American MARC format
MATER	Magnetic Tape Exchange for Terminological/Lexicographical Records
MEDLARS	Medical Literature Analysis and Retrieval System (US National Library of Medicine
MEDLINE	MEDLARS On-Line
MEKOFF-2	A MARC format in use in the CMEA countries
MeSH	Medical Subject Headings
MINISIS	Minicomputer software package based on ISIS
MONOCLE	Mise en Ordinateur d'une Notice Catalographique de Livre
MRDF	Machine-Readable Data File
MSUS/PALS	Minnesota State University System: Project for Automated Library Systems
NACO	Name Authority Co-operative Project
NAL	National Agricultural Library
NAM	Network Access Machine
NASA	National Aeronautics and Space Administration
NFAIS	National Federation of Abstracting and Indexing Services
NIH	National Institutes of Health
NLM	National Library of Medicine
NOTIS/LUIS	Northwestern Online Total Integrated Library System/Library User Information Service

NTIS	National Technical Information Service (US Department of Commerce)
OCLC	Online Computer Library Center
OCR	Optical Character Recognition
OECD	Organization for Economic Co-operation and Development
OL'SAM	Online Database Search Assistance Machine
PAD	Packet Assembler/Disassembler
PADIS	Pan African Documentation and Information System
PDI	Picture Description Instruction
POPIN	Population Information Network
POPLINE	Population Information Online
RAK	Regeln für alphabetische Katalogisierung
RLG	Research Libraries Group
RLIN	Research Libraries Information Network
RM	Reference Manual for Machine-Readable Bibliographic Descriptions
ROM	Read Only Memory
SCOPAL	Standing Conference of Pacific Libraries
SDI	Selective Dissemination of Information
SERIDU	Système Educatif de Recherche d'Information Documentation de l'ULB
TITUS	Textile Information Treatment User's Service
TLS	Total Library System
TOXLINE	Toxicology Information On-line
TSCA	Toxic Substances Control Act
UBC	Universal Bibliographic Control
UDC	Universal Decimal Classification
ULISYS	Universal Library System
UNBIS	United Nations Bibliographic Information System
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIBID	UNISIST International Centre for Bibliographic Descriptions
UNIMARC	Universal MARC Format
UNISIST	Programme of International Co-operation in Scientific and Technical Information
USERKIT	User Search Enhancement and Retrieval Kit for Interfacing to Terminals
VINITI	All-Union Institute of Scientific and Technical Information (USSR)
WEECN	Women's Educational Equity Communications Network
WLN	Washington Library Network